Bank Signalling, Risk of Runs, and the Informational Impacts of Regulations*

Kebin Ma[†] Tamas Vadasz[‡]

October 30, 2017

Preliminary version - Please do not cite or distribute without the author's permission

Download latest version

Abstract

Banks can take costly actions (such as higher capitalization, liquidity holding, and advanced risk management) to fend off runs. While such actions directly affect bank risks, they also carry informational content as signals of the banks' fundamentals. A separating equilibrium due to such signalling, however, involves two types of inefficiency: the high type chooses excessively costly signals, whereas the low type is vulnerable to runs. This provides a novel rationale for financial regulations: by restricting banks' actions, regulators can maintain a pooling equilibrium where the cross-subsidy among types promotes financial stability. We build a theoretical model to illustrate the point and also obtain supporting evidence from the US capital and liquidity regulations.

Keywords: Financial Regulation, Signalling, Bank Runs, Global Games

JEL Classification: G01, G11, G21

*The authors thank John Thanassoulis, Ansgar Walther, Lucy White, Alessandro Pavan, Adam Zawadowski for their insightful comments and discussion. Any remaining errors are our own.

[†]Warwick Business School, kebin.ma@wbs.ac.uk

[‡]Warwick Business School, t.vadasz@warwick.ac.uk

1 Introduction

Financial institutions often make use of costly actions to communicate private information about their fundamentals.[1] A particularly important class of such signalling devices appears to be the bank's quantitative risk management choices, such as the amount of capital or high-quality liquid assets, whose primary goal is to tackle solvency risk and liquidity risk respectively.

Based on this observation, we examine in this paper a largely overlooked aspect of such micro-prudential regulation: its impact on the value of information conveyed by risk management actions as a signalling device, and in turn, the incentives to engage in discretionary risk management. In a no-regulation environment, a separating equilibrium due to signalling involves two types of inefficiency: the high-type chooses excessively costly signals, whereas the low type is being revealed and becomes vulnerable to runs. In this sense, the private information creation can bear potential social inefficiency. We show, that a minimum quantitative regulation can squeeze out separating equilibrium and enforce pooling by making it more difficult to signal private information, thereby eliminating the related inefficiencies. This view of financial regulations fundamentally differ from the traditional ones that emphasize financial regulations' role in mitigating moral hazard of containing potential negative externalities on the real economy.[2]

This provides a novel perspective on financial regulation: by eliminating a way how markets create information, regulators create ignorance, which is efficient, as it leads to greater financial stability and higher social welfare. Our new perspective can also explain market participants' reaction to the introduction of a new regulation: if the new regime is sufficient in inducing pooling, institutions initially not constrained by the to-be introduced quantitative regulation, optimally decrease their level of risk management towards the new regulatory limit, which now serves as a focal point. The model also emphasizes a latent link between microprudential and macroprudential perspective: regulating certain individual entities changes the prevailing equilibrium, thereby the behaviour of other market participants, affecting the stability of the system as a whole.

Our model combines signalling with a stylized bank-run game, where the unique equilibrium of the coordination problem of creditors is determined by global games techniques. We assume that the strength of the bank is parameterized by two distinct fundamental variables: the bank's innate and its financial fundmental. While bank insiders have private information regarding the bank's innate ability to effectively make use of costly risk management tools to fend off runs, a lack of common knowledge regarding the financial

[1]A classic example would be banks maintaining high dividend payouts and executive compensation during the crisis, in the endeavour to convince the market of their relatively strong financial positions.

[2]For example, minimum capital regulation is often justified to correct moral hazard and risk-taking incentives of shareholders, while the recent introduction of quantitative liquidity regulation is motivated by decreasing reliance on 'public liquidity' and building up sufficient private cushions to withhold liquidity shocks.

fundamental drives the global game equilibrium selection in the second stage of the game.

This model can be solved similarly to a conventional signalling game, with an important difference: the receivers' (creditors) unique aggregate responses to any on- and off-equilibrium action by the sender are determined by global-games techniques. Our methodological contribution is to develop a novel technique to analyse a global game embedded into a signalling game in a way which also facilitates welfare analysis. We illustrate our approach with a linear regime switching function which leads to a closed-form solution, and generalize to a larger class of models satisfying a single-crossing property.

The first main result of the paper shows the existence of a separating equilibrium in which a high-type sender sends an excessively high signal. We show that the existence of this equilibrium, as well as the magnitude of inefficiency is inherently linked to the precision of receivers' noisy private signals, with higher precision leading to larger inefficiencies. Intuitively, higher precision leads to a more correlated aggregate behaviour in any states of the world, in turn leading to more pronounced aggregate responses to (perceived) changes of unknown parameters, thereby making the low-type more aggressively mimic the high-type.

Second, we show that a regulator can eliminate *high signals* by setting a *minimum* threshold on the bank's risk management action. Under such regulations, the minimax payoff for the low-type (i.e. the payoff she could get irrespectively of the other type's behaviour) decreases, which in turn increases the critical policy that is required for the high-type to maintain a separating equilibrium. For a sufficiently restrictive regulation, the required intervention is too high, and it will no longer be incentive-compatible for the high-type to maintain the separating equilibrium. This *critical* regulatory minimum also changes with precision: when precision is high, the separating signal is also high, therefore a lower minimum regulation is sufficient. This finding can explain the emergence of signalling in turbulent times (characterized by low precision of observation of fundamentals): a certain regulation which is just sufficient to maintain pooling during normal times might not be able to prevent wasteful signalling during turbulent times.

We perform a preliminary welfare analysis tailored specifically to the context of financial regulations. We show that a minimum ratio regulation can increase ex-ante welfare by squeezing out separating equilibrium. Indeed, a separating equilibrium in the model leads to two types of inefficiency: the high-type chooses a signal that is excessively high and costly, whereas the low-type is identified as weak and becomes vulnerable—resulting in more runs and greater financial instability. In contrast, in a pooling equilibrium, the high-type will cross-subsidise the low-type, and the economy can feature greater financial stability as well as a reduction of costly signaling. In this sense, financial regulations reduce the information available in private markets, and the resulting ignorance is efficient. The ex-ante improvement in expected profits implies that it can be incentive-compatible for banks to accept financial regulations.

Our model leads to testable empirical predictions. If financial regulations do squeeze out separating equilibrium, we would expect relatively high dispersion of risk management measures among banks before the introduction of pertinent financial regulations, and a clustering of observations after the introduction of the regulations. We test this hypothesis on two data sets: cash holding of US Bank holding companies (BHC's), using a difference-in-difference method which exploits the recent introduction of Liquidity Coverage Ratio (LCR) regulation, as well as changes in capital ratios around the introduction of Basel I regulatory capital regime. We find two distinct patterns, both consistent with the predictions of our theory: first, the dispersion of cash ratios for BHC's subject to the new regulation have fallen significantly more sharply than those which were not subject to the new regulation. This is consistent with a successful elimination of separating equilibrium. For the capital regulation, we find a sudden increase of the number of institutions with large equity ratios, which might be the result of an insufficient regulatory minimum, being unable to squeeze out, but boosting the signals required to maintain separating equilibrium.

Related literature. The idea that simple risk management measures such as capital or liquidity can signal private information above and beyond the fact that higher values can mechanically protect the bank against shocks is not new in the literature. For example, Hughes and Mester (1998) writes: *"Since financial capital constitutes the bank's own bet on its management of risk, it conveys a credible signal to depositors of the resources allocated to preserving capital and insuring the safety of their deposit"*. The signalling role of capital is also well recognized generally in the corporate finance literature, albeit with somewhat inconclusive predictions (Ross (1977),Brealey et al. (1977) and Harris and Raviv (1991)). In Malherbe (2014), liquidity can be interpreted as a signal of the underlying reason of asset sale by banks, thereby a higher liquidity might increase adverse selection on asset markets. In other papers where asymmetric information concerns the quality of assets, banks' are sending credible signals either trough proper security design (Nicolo and Pelizzon (2008)) or by retention (He (2009)). An extensive literature in accounting surveys the signalling role of loan loss provisioning (LLP), with many papers arguing that higher LLP credibly signals a prudent risk management, and management's intention to resolve problem debt situations[3].

Our model predicts positive relationship between the level of risk management measures and the value of the bank in case of first-best as well as whenever separating equilibrium still prevails on the markets. In case of capital for example, this is consistent with Mehran and Thakor (2010), who present an elegant theory and strong empirical support for a positive cross-sectional correlation between bank capital and market value. In their model, increasing capital has two effects: it increases probability of survival and in turn,

[3]Some early contributions are: Beaver and Engel (1996), Scholes et al. (1990), Grammatikos and Saunders (1990), Griffin and Wallach (1991)

incentives to monitor (direct effect), while increased loan monitoring enhances the value of the portfolio (indirect effect). The overall impact of the two effects is that banks with lower monitoring costs will have higher marginal benefit on capital, and in turn, find it optimal to hold more of them.

The methodology of our paper is most related to Angeletos et al. (2006), and Angeletos and Pavan (2013). In their pioneering work, the authors consider a perfectly informed policy maker (sender) who tries to defend a regime with possible policy intervention and show that once the signalling effect of the policy intervention is taken into consideration, multiple equilibria will re-surface in a global-game setting due to the endogeneity of attackers' (receivers') information set. The specific form of multiplicity in the (semi-) separating equilibrium arises due to the fact that there is no uncertainty regarding the regime outcome from the sender's perspective. As a consequence, any positive policy intervention signals the survival of the regime, which makes 'no attack' a dominant strategy, and the global game is played out over a truncated posterior distribution on the range of fundamentals when intervention does not occur.

In contrast to this work, as well as a growing literature on persuasion with multiple receivers (Inostroza and Pavan (2017), Goldstein and Huang (2016)) the sender in our model only imperfectly observes the fundamentals. In the context of banking, insiders such as bank equity holders or managers (i) can have information advantage over their creditors regarding the bank's resilience to shocks but (ii) still face uncertainty regarding the fate of the bank. Apart from being more realistic in the context of banking risk management, the modelling role of residual uncertainty on the sender's side is crucial: despite her information advantage, this additional uncertainty keeps the sender uncertain regarding the fate of the regime, therefore policy intervention cannot make even the highest type bank completely 'run-proof', although changes incentives to run.

The paper is also related to the large literature on bank runs. Since the seminal contribution of Diamond and Dybvig (1983) it is well known that liquidity transformation makes the banks vulnerable to runs driven by agents' self-fulfilling beliefs regarding the behaviour of other agents. In the more recent follow-up literature, global games theory[4] has been routinely used to resolve the equilibrium selection problem in the Diamond-Dybvig framework (Goldstein and Pauzner (2005)). It has also been pointed out that liquidity, as well as capital – ceteris paribus – can serve as a buffer, thereby dampen the probability of distress and increase financial stability (Diamond and Rajan, 2000, Diamond and Kashyap, 2016).

The paper is organized as follows: we introduce our model in Section 2. In Section 3, we analytically solve the model for a no-regulation equilibrium with stylized functional forms. Section 4 analyses the impact of a minimum quantitative regulation and discusses the most important welfare trade-offs. Section 5 provides some empirical insights. Finally,

[4]Carlsson and Van Damme (1993) and Morris and Shin (1998)

we generalize some of our results to a larger class of functional forms in Appendices.

2 Model setup

We consider a game played by two groups of players: a bank's management and the bank's creditors. The creditors hold demandable claims and simultaneously decide whether to run on the bank or not, while the management is incentivized to defend the bank from runs by implementing costly risk management.

The bank's fundamental is characterized by two variables: $\theta_1 \in \{\theta_1^L, \theta_1^H\}$ can be interpreted as the bank's risk management skill, while $\theta_2 \in \mathbb{R}$ is the financial fundamental. We assume that the two fundamentals are independent with the following prior distributions: $Prob(\theta_1 = \theta_1^L) = p_L$ and $Prob(\theta_1 = \theta_1^H) = p_H$ with $p_L + p_H = 1$, while θ_2 is uniform on $[\overline{\theta}_2 - 2\eta, \overline{\theta}_2]$. We write $\boldsymbol{\theta} := \{\theta_1, \theta_2\}$ to simplify notation.

In period 1 the management privately observes the realization of θ_1 which is the bank's Harsanyi-type, and chooses a real-valued, costly risk-management action $s \in (0, +\infty)$ to enhance the bank's ability to survive runs. The management's strategy, therefore, specifies a choice of s for each possible realization of θ_1. The action s directly influences the ability to survive runs and also serves as an informative public signal of the bank's type.

The bank fails if sufficiently many creditors decide to run. In particular, the bank's failure can be represented by a continuous, differentiable, real-valued regime switching function $\mathcal{R}(\boldsymbol{\theta}, s, \alpha)$, with the bank failure occurring whenever $\mathcal{R}(\boldsymbol{\theta}, s, \alpha) < 0$, where α is the size of the aggregated attack. We assume that $\mathcal{R}(\boldsymbol{\theta}, s, \alpha)$ satisfy natural requirements[5] $\mathcal{R}_{\theta_2} > 0$, $\mathcal{R}_{\theta_1} > 0$, $\mathcal{R}_{\alpha} < 0$, $\mathcal{R}_s > 0$ and $\mathcal{R}_{s\theta_1} > 0$. For the main part of the paper, we solve the model analytically for the following functional form.

$$\mathcal{R} = \theta_1 s + \theta_2 - \alpha \tag{1}$$

So the fundamental θ_1 is a measure of how effective the bank's pre-emptive intervention can be in avoiding bankruptcy.[6]

The management's payoff is as follows.

$$U(\boldsymbol{\theta}, s, \alpha) = \begin{cases} k - c \cdot s & \text{if} \quad \mathcal{R}(\boldsymbol{\theta}, s, \alpha) \geq 0 \\ 0 & \text{if} \quad \mathcal{R}(\boldsymbol{\theta}, s, \alpha) < 0 \end{cases}$$

where k is the benefit of defending the regime, c is the cost of policy action s, and the

[5]The bank's survival is more likely if any of the fundamentals or the risk management action are higher, less likely if the mass of creditors who run is higher. Any fixed s is more helpful for a higher-quality bank.

[6]For example, if a higher s represents the use of more advanced risk-managment modelling, a high θ_1 would be the human resource or IT infrastructure required in the implementation. Or, if a higher s represents a higher capital level, a high θ_1 would indicate the low agency cost to raise external capital.

payoff conditional on failure is normalized to zero.[7] The parameters c and k are exogenous constants.

In period 2, a unit mass of creditors observe the public signal s and a private, noisy signal of the bank's financial fundamental θ_2. In particular, on top of perfectly observing s, creditor $i \in [0, 1]$ observes $x_i = \theta_2 + \sigma\epsilon_i$, where ϵ_i are independently and identically distributed uniformly on $[-1, 1]$. Based on this information, the creditors simultaneously decide whether to run on the bank. The payoff from running is normalized to $0 < t < 1$, while the payoff from not running is as follows.

$$u(\boldsymbol{\theta}, s, \alpha) = \begin{cases} 1 & \text{if } \mathcal{R}(\boldsymbol{\theta}, s, \alpha) \geq 0 \\ 0 & \text{if } \mathcal{R}(\boldsymbol{\theta}, s, \alpha) < 0 \end{cases}$$

The timeline is summarized in Figure 1:

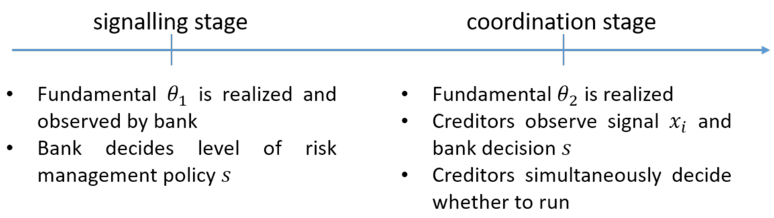


Figure 1: Timeline

Note that this information structure arises naturally in a dynamic settings: bank's risk management decisions might well be based on private information, but precede in time the actual realization of economic fundamentals (for example the default rate on credit portfolio or return realization on long-term assets) and the decision of depositors whether to run on the bank.

The simple functional forms for the regime switching function as well as for the payoff functions are selected to sharpen the intuition and emphasize the interaction between the signalling and the coordination components as clearly as possible.[8] In Appendix B we analyse some of the consequences of these modelling choices and generalize the results to more general class of functions.

[7]One interpretation of the payoff structure is that k captures the debt-like claims that the executives hold such as salary and pension compensation, while the zero payoff in the case of bank failure reflects individuals' limited liability to the firm they manage.

[8]Indeed, the model can be straightforwardly recast as a backbone 'regime switching' game where asymmetrically informed atomistic players (creditors) play a game with strategic complementarities whether to attack (stay) or not (withdraw) a regime (bank), of which survival depends on its fundamentals ($\boldsymbol{\theta}$), actions (s), and the mass of atomistic players attacking the regime (α).

3 Equilibrium analysis

The equilibrium concept for this and the next chapter is Perfect Bayesian Equilibrium. Let $s(\theta_1)$ denote the strategy of banks' management, i.e. the policy chosen by each type, $a_i(x_i, s)$ denote the action of an agent receiving private signal x_i and observing policy action s, α denote the aggregate attack size. We define equilibrium as follows.

Definition 1 *An equilibrium of the signalling-global game consist of (1) a strategy* $s(\theta_1) : \Theta_1 \to [\underline{s}, \infty)$ *for bank, (2) a strategy for creditors,* $a_i(x_i, s) : \mathbb{R} \times [\underline{s}, \infty) \to \{0, 1\}$*; (3) A posterior belief on* $\boldsymbol{\theta}$ *for creditors upon observing* $\{x_i, s\}$ *and (4) an aggregate attack function* α*, such that*
(1) The bank and the creditors are sequentially rational: $s(\theta_1)$ *maximizes utility:*

$$s(\theta_1) = \arg\max_s \mathbb{E}_{\theta_2} U(\boldsymbol{\theta}, s, \alpha(\boldsymbol{\theta}, s)|\theta_1)$$

and $a_i(x_i, s)$ *maximizes agents utility*

$$a_i = \arg\max_{a_i} \mathbb{E}_{\boldsymbol{\theta}} u(\boldsymbol{\theta}, s, a_i|x_i, s)$$

where aggregate attack α *is calculated from individual decisions*

$$\alpha(\boldsymbol{\theta}, s) = \int_0^1 a_i di$$

(2) The expectations are taken according to the bank's (agents') posteriors $\mu(\theta_2|\theta_1)$ *(and* $(\phi(\boldsymbol{\theta}|x_i, s))$*) respectively, which are calculated using Bayes' rule whenever possible.*

We solve the model with backward induction: first solve the coordination game for any given policy level s, then determine sender's optimal choice given 2nd stage equilibrium.

3.1 Symmetric information benchmark

To set up a standard benchmark, we start by solving a version of the game where θ_1 is observed by creditors as well. Without information asymmetry regarding θ_1, there is no signalling role of policy action. Second stage equilibrium is calculated using standard global games techniques with exogenously given s. Lemma 1 summarizes the result.

Lemma 1 *When parameter* θ_1 *is perfectly observed by the creditors, the unique equilibrium that survives iterated elimination of strictly dominated strategies is characterized by* *two thresholds*[9]

$$\hat{x} = t - \theta_1 s + 2\sigma t - \sigma \tag{2}$$

$$\hat{\theta}_2 = t - \theta_1 s \tag{3}$$

such that creditor i runs if and only if $x_i < \hat{x}$, and the bank fails if and only if $\theta_2 < \hat{\theta}_2$.

Proof. See Appendix ■

According to Lemma 1, the equilibrium of the subgame is characterized by a pair $\{\hat{\theta}_2, \hat{x}\}$ which jointly solves two equations: (i) a creditor who receives signal $\hat{x}$ is just indifferent between RUN and WAIT, and (ii) the bank just fails at $\hat{\theta}_2$. The proof in Appendix formulates these two conditions and solves for the two equilibrium values. Both thresholds are decreasing in type θ_1 and in action s, which implies that higher fundamental, as well as higher intervention makes survival of the regime more likely.

Next, we solve for optimal risk management, given that the bank anticipates correctly the equilibrium in the second stage. For any choice of s, the equilibrium quantities $\{\hat{\theta}_2, \hat{x}\}$ determine the mass of agents who run on the bank, and in turn, the probability of survival. The objective function therefore can be expressed as a function of the exogenous type θ_1 and the banks' endogenous intervention s. Let $\rho(\cdot)$ denote the probability of survival:

$$\rho(\theta_1, s) = Pr[\theta_2 < \hat{\theta}_2 | \theta_1, s] = \frac{1}{2\eta}\left(\overline{\theta}_2 - \hat{\theta}_2(\theta_1, s)\right)$$

Then we can write the expected profit for any given signal s as

$$\pi(\theta_1, s) = \int_{\hat{\theta}_2}^{\overline{\theta}_2} (k - c \cdot s)\, \mu(\cdot) d\theta_2 = \rho(\theta_1, s)(k - c \cdot s)$$

where $\mu(\cdot)$ denotes the prior on θ_2. Optimal intervention trades off cost of intervention with an increased probability of survival. The first-order condition

$$\frac{\partial \pi}{\partial s} = \frac{\partial \rho}{\partial s}(k - c \cdot s) - \rho c = \frac{1}{2\eta}\left(\theta_1(k - c \cdot s) - c\overline{\theta}_2 + c(t - \theta_1 s)\right) = 0$$

implies the optimal intervention $s^\star(\theta_1)$ and the associated optimal profit $\pi^\star(\theta_1)$ as

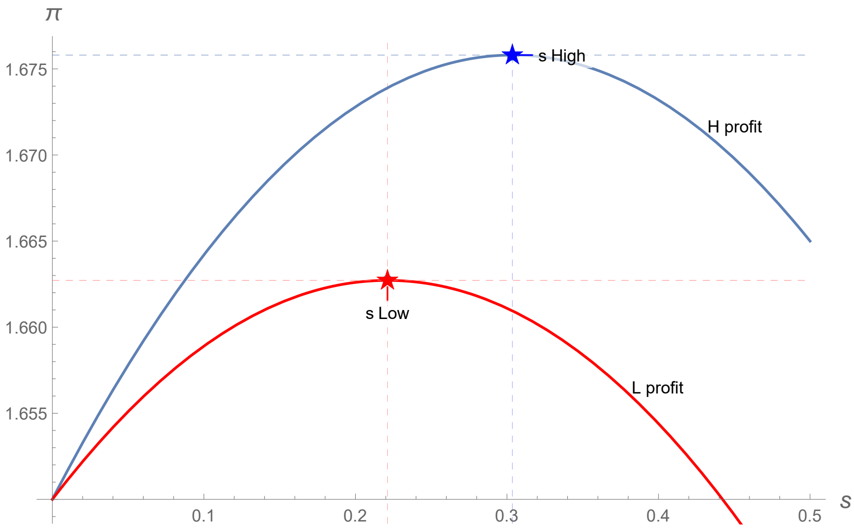
$$s^\star(\theta_1) = \frac{1}{2}\left(\frac{k}{c} - \frac{\overline{\theta}_2 - t}{\theta_1}\right) \tag{4}$$

$$\pi^\star(\theta_1) = \frac{\left(c\left(\overline{\theta}_2 - t\right) + \theta_1 k\right)^2}{4c\theta_1} \tag{5}$$

The symmetric information benchmark equilibrium is summarized in Lemma 2:

[9] Following the literature, we will refer to (variants of) $\hat{\theta}_2$ as 'fundamental threshold', while $\hat{x}$ as 'strategic threshold'.

Figure 2: Expected profit as a function of intervention



Expected profit is increasing in type, and the optimal intervention satisfies $s_L^\star < s_H^\star$

Lemma 2 *When type* θ_1 *is public information, a sender of type* θ_1 *optimally sets* $s^\star(\theta_1)$ *according to Equation* (4) *and obtains payoff* $\pi^\star(\theta_1)$ *as in Equation* (5). *In equilibrium, provided* $\frac{k}{c} > \frac{\bar{\theta}_2 - t}{\theta_1}$, *both the optimal intervention and profit increase in parameter* θ_1.

Proof. See Appendix ■

Figure 2 illustrates expected profits as a function of signal s, and optimal signals for high (H) and low (L) type banks. First-best interventions are higher for the H-type, which is a direct consequence of the higher marginal benefit of action. The condition in the Lemma is a necessary and sufficient condition for the positivity of the optimal intervention, which we can assume for the problem to be interesting.

3.2 Separating equilibrium

Moving towards analysing asymmetric information, we first characterize separating equilibrium. In any pure-strategy separating equilibrium the two types send different signals

$$s(\theta_1) := \begin{cases} s_L & \text{if} \quad \theta_1 = \theta_1^L \\ s_H & \text{if} \quad \theta_1 = \theta_1^H \end{cases}$$

and the chosen action reveals the type perfectly to creditors. Separating equilibrium can be maintained if no player has incentives to deviate. Before formalizing the incentive compatibility constraints, the first step is to compute payoffs off-the-equilibrium path.

3.2.1 Off-equilibrium payoffs

First, consider the case where a sender of type L chooses an off-equilibrium action and mimics type H. The receivers - believing that they are facing a H-type sender - behave as if they were facing H-type with certainty. Therefore, their optimal response given these beliefs is described by the fundamental threshold defined in Lemma 1 for type H:

$$\hat{x}^H = t - \theta_1^H s^H + 2\sigma t - \sigma$$

This implies that the mass of agents who would run upon any realization of θ_2 is exactly the same as if it is under type H, that is, $\alpha(\hat{x}^H, \theta_2)$[10]. The off-equilibrium fundamental threshold for the deviating type, denoted by $\hat{\theta}_2^{L.H}$ is the value of θ_2 which solves

$$\theta_1^L s^H + \theta_2 - \alpha(\hat{x}^H, \theta_2) = 0$$

which implies, after substituting the expression for α and rearranging

$$\hat{\theta}_2^{L.H} = \frac{\hat{x}^H + \sigma - \sigma s^H \theta_1^L}{2\sigma + 1}$$

Substituting $\hat{x}^H$ and defining the type difference $\Delta\theta_1 = (\theta_1^H - \theta_1^L)$ we obtain

$$\hat{\theta}_2^{L.H} = \frac{t(1+2\sigma) - \theta_1^H s^H - 2\sigma s^H \theta_1^L}{2\sigma + 1} = \hat{\theta}_2^H + \frac{2\sigma s^H \Delta\theta_1}{1+2\sigma} = \hat{\theta}_2^L - \frac{s^H \Delta\theta_1}{1+2\sigma} \tag{6}$$

Analogously, by replacing indices but keeping the definition of $\Delta\theta_1 = (\theta_1^H - \theta_1^L)$ fixed, it is possible to define off-equilibrium thresholds for the case when H mimics L[11]

$$\hat{\theta}_2^{H.L} = \hat{\theta}_2^L - \frac{2\sigma s^L \Delta\theta_1}{1+2\sigma} = \hat{\theta}_2^H + \frac{s^L \Delta\theta_1}{1+2\sigma}$$

By mimicking the other type's action in an off-the-equilibrium path of a candidate separating equilibrium, the sender can influence the behaviour of receivers and induce them to behave according to the strategy what they would follow under the other type. However, he cannot achieve the same fundamental threshold, since the true type enters directly into the regime change function $\mathcal{R}$, which determines the threshold.

Before deriving equilibrium, we introduce an alternative interpretation of off-equilibrium thresholds.For any given (not necessarily equilibrium) s, the functions $\hat{\theta}_2^{L.H}(s)$ and $\hat{\theta}_2^{H.L}(s)$ can be understood as failure thresholds for the given type *if his creditors believe it to be the other type.* These functions define an additive decomposition of the difference between the two types' *first-best* fundamental thresholds for a given fixed s. For example, using

[10]The expression for that can be found in the proof of Lemma 1, Appendix A.1.

[11]As will be clear in the equilibrium analysis, H would never want to mimic L. Yet, he might find it optimal to go 'off-path', in which case his payoffs are characterized by $\hat{\theta}_2^{H.L}$

Equation 6, we can write

$$\hat{\theta}_2^L(s) - \hat{\theta}_2^H(s) = \underbrace{\frac{2\sigma s\Delta\theta_1}{1+2\sigma}}_{\text{Direct effect}} + \underbrace{\frac{s\Delta\theta_1}{1+2\sigma}}_{\text{Indirect effect}} \tag{7}$$

Equation 7 decomposes the difference between first-best fundamental thresholds into a sum of a *direct effect*, attributable to the fundamental difference between types L and H, and an *indirect effect*, which is solely due to creditors' beliefs. As we will show in the next section formally, the larger the indirect effect is, the more can a low-type potentially benefit from mimicking the high type, and similarly, the larger is the potential loss for a high type for not being able to distinguish himself from a low type. Note that the indirect effect decreases in the noise σ.

Expected profit off-the-equilibrium path

If type L is believed to be[12] type H at *some* signal s, she obtains expected profit

$$\pi^{\text{off}}(s, \underbrace{\theta_1^L}_{\text{true}}, \underbrace{\theta_1^H}_{\text{perceived}}) = \int_{\hat{\theta}_2^{L.H}}^{\overline{\theta}_2} (k - c\cdot s)\,\mu(\cdot)d\theta_2 = \rho^{\text{off}}(s, \theta_1^L, \theta_1^H)(k - c\cdot s) \tag{8}$$

where analogously to previous notation we define the probability of survival off-the equilibrium path when type L mimics type H at *some* s as:

$$\rho^{\text{off}}(s, \theta_1^L, \theta_1^H) = \frac{1}{2\eta}\left(\overline{\theta}_2 - \hat{\theta}_2^{L.H}\right)$$

Optimal off-equilibrium action is again derived from the first-order condition

$$\frac{\partial\pi^{\text{off}}(s, \theta_1^L, \theta_1^H)}{\partial s} = \frac{\partial\rho^{\text{off}}(s, \theta_1^L, \theta_1^H)}{\partial s}(k - c\cdot s) - \rho^{\text{off}}(s, \theta_1^L, \theta_1^H)\cdot c$$

which implies the optimum off-equilibrium intervention

$$s^{\star}_{L.H} = \frac{1}{2}\left(\frac{k}{c} + \frac{(1+2\sigma)(t-\overline{\theta}_2)}{2\sigma\theta_1^L + \theta_1^H}\right) = 0 \tag{9}$$

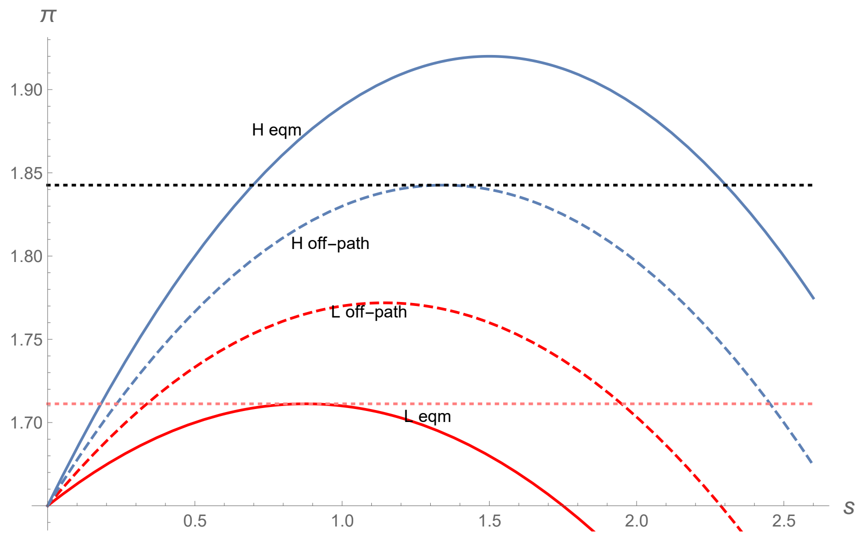
and off-equilibrium optimal profits:

$$\pi^{\star}_{L.H} := \pi^{\text{off}}(s^{\star}_{L.H}, \theta_1^L, \theta_1^H) = \frac{\left(-c(2\sigma+1)\left(t-\overline{\theta}_2\right) + k(2\sigma\theta_1^L + \theta_1^H)\right)^2}{4c(2\sigma+1)\left(2\sigma\theta_1^L + \theta_1^H\right)} \tag{10}$$

Analogous expressions can be derived for $s^{\star}_{H.L}$ and $\pi^{\star}_{H.L}$. It is straightforward to show

[12] that happens if she successfully mimics type H

Figure 3: On (solid) and off-path (dashed) equilibrium payoffs



that the following relationships hold:

$$s^\star_L < s^\star_{L.H} \qquad \text{and} \qquad s^\star_H > s^\star_{H.L}$$
$$\pi^{eq}(s, \theta_1^L) < \pi^{\text{off}}(s, \theta_1^L, \theta_1^H) \qquad \text{and} \qquad \pi^{eq}(s, \theta_1^H) > \pi^{\text{off}}(s, \theta_1^H, \theta_1^L) \qquad \forall s$$

Off-equilibrium payoffs and optimal actions are critical in analysing the existence of equilibrium. In particular, the profit $\pi^\star_{H.L}$ is H-type's minimax payoff: even with the most adverse beliefs of creditors (if all believe he is of bad type), he can obtain payoff at least $\pi^\star_{H.L}$. Therefore, in any proposed equilibrium, type H's payoff must exceed $\pi^\star_{H.L}$. Note that the low type's (L) minimax payoff is $\pi^\star_L$.

Figure 3 illustrates on- and off-equilibrium payoffs as a function of an arbitrary policy intervention s. In the next section we establish which values of s can be maintained as separating equilibrium.

3.2.2 Characterization of equilibrium

Equilibrium requires that no types have incentives to deviate from equilibrium actions $\{s_L, s_H\}$. First, note that in any separating equilibrium types are revealed, so L-type will find it optimal to set $s_L = s^\star_L$, and obtain profit $\pi^\star_L$. A separating equilibrium in which the high-type sets some value s_H and the low type sets her optimum value $s^\star_L$ can be maintained if and only if

$$\pi^{eq}(s^\star_L, \theta_1^L) \geq \pi^{off}(s^H, \theta_1^L, \theta_1^H) \qquad (IC_L)$$
$$\pi^{eq}(s_H, \theta_1^H) \geq \pi^{off}(s^\star_{H.L}, \theta_1^H, \theta_1^L) \qquad (IC_H)$$

Let's denote s_L^{cri} the value of s_H which solves $[IC_L]$, that is, the value of signal at which the L-type is just indifferent between mimicking the high type, or setting $s_L^\star$ and obtaining his minimax profit. This is the value of s which solves

$$\rho^{eq}(s_L^\star, \theta_1^L)(k - c \cdot s_L^\star) \geq \rho^{off}(s, \theta_1^L, \theta_1^H)(k - c \cdot s)$$

The RHS is a quadratic function with a negative coefficient of the quadratic term and with maximum value exceeding the constant on the LHS, so the corresponding equality has two solutions $\left(s_{L,1}^{cri} < s_{L,2}^{cri}\right)$. Incentive compatibility requires that $s_H \notin [s_{L,1}^{cri}, s_{L,2}^{cri}]$, otherwise L would have an incentive to mimic H. It is yet to be proven formally that in this specific case $\pi^{eq}(s_{L.1}^{cri}, \theta_1^H) < \pi^{eq}(s_{L.2}^{cri}, \theta_1^H)$ which implies that the individually rational choice for the good type is to send high signal, and the level which can maintain separating equilibrium must fulfil $s_H \geq s_{L.2}^{cri}$. [13]

Similarly, define s_H^{cri} to be the critical s which is incentive-compatible for type H and solves $[IC_H]$. This is the level of intervention at which the profit of a good type in separating equilibrium is at least as much as its best profit if it is believed to be low-type. This latter utility is the good types' minimax payoff - irrespectively to player's beliefs he can always achieve $\pi_{H.L}^\star$ by setting the off-equilibrium profit-maximizing level of s.[14] Following the same formalism we have

$$\rho^{eq}(s_H, \theta_1^H)(k - c \cdot s_H) \geq \rho^{off}(s_{H.L}^\star, \theta_1^H, \theta_1^L)(k - c \cdot s_{H.L}^\star)$$

which by similar argument, has two solutions $s_{H.1}^{cri}$ and $s_{H.2}^{cri}$. We can characterize the existence of separating equilibrium in terms of the thresholds derived above as follows.[15]

1. Separating equilibrium exists and it restores the first best if and only if $s_{L.2}^{cri} \leq s_H^\star$.
2. Separating equilibrium exists and in this equilibrium the good type (H) sends inefficiently high signal if and only if $s_H^\star < s_{L.2}^{cri} \leq s_{H.2}^{cri}$.
3. Separating equilibrium does not exist if and only if $s_{H.2}^{cri} < s_{L.2}^{cri}$.

We derive closed analytical formulas for the critical values Appendix along with some limiting cases, which we will use in the following discussion. Theorem 1 establishes the link between the precision of private signals and the existence of separating equilibrium.

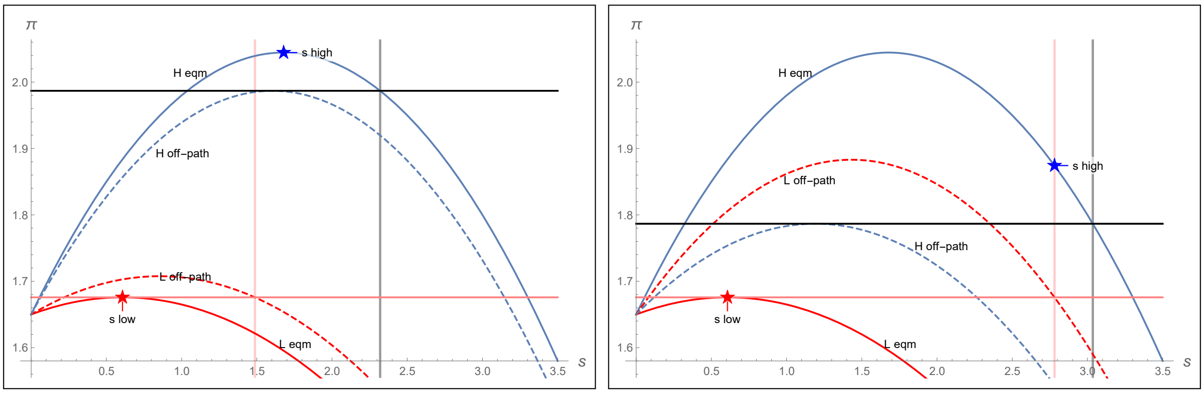
Theorem 1 *There exists an 'efficient' separating equilibrium in which signals coincide with the first-best if and only if the noise in creditors' private observation is sufficiently*

[13] We derive analytical expressions in Appendix .

[14] Note the difference: L-type's minimax payoff is $\pi_L^\star$ while type H's minimax payoff is $\pi_{H.L}^\star$ - at least under the specified beliefs.

[15] we will only focus on the 'upper' regions, because it is more important for our application. It is straightforward to extend the analysis to the lower part.

Figure 4: Separating Equilibrium



(a) The first-best is restored in Separating Equilibrium with high σ (low precision)

(b) High-type sends inefficiently high intervention with low σ (high precision)

large, that is, if and only if

$$\sigma \geq \overline{\sigma}$$

Whenever $\sigma < \overline{\sigma}$, and $c > \hat{c}$, there exists an 'inefficient' separating equilibrium in which the sender must choose a higher-than-the-first-best policy intervention, where $\hat{c}$ is defined as

$$\hat{c} = \sqrt{\frac{\Delta\theta_1 (k\theta_1^L)^2}{(3\theta_1^L + \theta_1^H)\left(\overline{\theta}_2 - t\right)}} \tag{11}$$

Whenever $c < \hat{c}$, there exists a lower boundary $\underline{\sigma}(c)$ such that separating equilibrium does not exists for every $\sigma < \underline{\sigma}(c)$.

Proof. See Appendix ■

The Theorem is illustrated in Figure 4. The precision of creditor's signal $(1/\sigma)$ has a critical role in determining which type of equilibrium can survive. If information is less precise, the potential benefit/loss from mimicking the other type decreases. This is because if type L mimics type H, the (off-the-equilibrium-path) strategic threshold $\hat{x}$ is pinned down according to the equilibrium of type H, but the fundamental threshold, which enters directly into the integral boundary of the expected profit, is not.[16] A higher noise pushes the realized fundamental threshold upwards, thereby decreasing the profit which can be obtained by mimicking type H. This, in turn, decreases $s_{L.2}^{cri}$, which approaches $s_L^\star$ as $\sigma \to \infty$. Since $\lim s_{L.2}^{cri} < s_H^\star$, due to continuity there exist an σ (denoted by $\overline{\sigma}$) at which $s_{L.2}^{cri}(\overline{\sigma}) = s_H^\star$. Consequently, if and only if the fundamental is observed with large enough noise $(\sigma > \overline{\sigma})$ is the separating equilibrium efficient (i.e. restores sender-optimal first-best). If the equilibrium is inefficient, the distortion increases as the noise becomes more precise.

[16] A way to think about this is: in the standard global game, the fundamental threshold is fixed when precision in varied, and the strategic threshold adapts to the changes. In contrast, when L mimics H, the strategic threshold is fixed and the realized fundamental threshold varies. When the strategic threshold of H shifts to the left due to increased noise, the fundamental threshold must shift to the left as well.

Intuition: the more precise receivers' private observation is, the more 'extreme' is the *effective strategy*, defined as the probability of attack for any realization of fundamental θ_2. In particular, with $\sigma \to 0$ the effective strategy converges to a limiting case where all agents attack if and only if $x_i \leq \hat{\theta}_1$, no agents attack otherwise. However, the more extreme is the effective strategy, the more important it is to 'get the other parameters right'. The aggregated *strategic error* by following a certain strategy which happens to be wrong is the largest, when the information which determines the strategy is the most precise. If the effective strategy is more flat, the differences of not knowing the other parameter (θ_1) correctly are 'smoothed out' by the relative flatness of the effective strategy.

3.3 Pooling equilibrium

In a pooling equilibrium the same signal is chosen by both types, which does not convey any information to the receivers. Equilibrium of the coordination stage is determined analogously to a standard global game with an important twist: given creditors' strategy $\hat{x}$, which must be the same under both types in a pooling equilibrium, the fundamental threshold for the two types will be different. This has to be taken into consideration by agents when calculating equilibrium strategies. In conclusion, any pooling equilibrium is characterized by a common strategic threshold for agents $\hat{x}$, and a distinct fundamental threshold for each types, $\hat{\theta}_2^L \neq \hat{\theta}_2^H$ in which (i) agents attack if and only if $x_i \leq \hat{x}$ and (ii) regime of type I (resp. H) fails if and only if $\theta_2 \leq \hat{\theta}_2^L$ (resp. $\leq \hat{\theta}_2^H$). The equations determining equilibrium of the global game change accordingly: (i) an agent who receives $\hat{x}$ should be indifferent between run and stay with the bank, given that banks of type L, H fails if and only if fundamental θ_2 is below the respective threshold and the (posterior) beliefs are (p_L, p_H), and (ii) bank of type $L(H)$ fails exactly at $\hat{\theta}_2^L(\hat{\theta}_2^H)$ if creditors run if and only if $x_i < \hat{x}$. Pooling equilibrium thresholds are characterized by Lemma 3.

Lemma 3 *In any pooling equilibrium in which both types send the same signal - denoted by s_p - the equilibrium of stage 2 (coordination game) is characterized by fundamental thresholds $\hat{\theta}_2^L$ and $\hat{\theta}_2^H$ and strategic threshold $\hat{x}$ where*

$$\hat{\theta}_2^L = t - \frac{s_p\overline{\theta}_1}{1+2\sigma} - \frac{2\sigma s_p \theta_1^L}{1+2\sigma}$$
$$\hat{\theta}_2^H = t - \frac{s_p\overline{\theta}_1}{1+2\sigma} - \frac{2\sigma s_p \theta_1^H}{1+2\sigma}$$

and

$$\hat{x} = 2\sigma t - \sigma + p_L\hat{\theta}_2^L + p_H\hat{\theta}_2^H \tag{12}$$

Proof. See Appendix ■

Corollary 1 *The average fundamental threshold in any pooling equilibrium is a linear function of the average type* $\overline{\theta}_1 = p_L\theta_1^L + p_H\theta_1^H$

$$\overline{\theta}_2^P = t - \frac{s(p_L\theta_1^L) + p_H\theta_1^H)}{1+2\sigma} - p_L\frac{2\sigma s\theta_1^L}{1+2\sigma} - p_H\frac{2\sigma s\theta_1^H}{1+2\sigma} = t - \overline{\theta_1}s$$

This result will be used to formalize ex-ante welfare. The pooling equilibrium thresholds can be rewritten as

$$\hat{\theta}_2^L = \hat{\theta}_2^{L,FI} - \frac{sp_H\Delta\theta_1}{1+2\sigma}$$
$$\hat{\theta}_2^H = \hat{\theta}_2^{H,FI} + \frac{sp_L\Delta\theta_1}{1+2\sigma}$$

where FI stands for the full-information threshold. Some consequences can be seen immediately. First, as the noise increases, we approach the full-information thresholds, thereby the risk-sharing effect of pooling decreases. The intuition for this is similar to that of separating equilibrium thresholds: with higher noise, fundamental thresholds become less responsive to other parameters of the game. Second, as the ex-ante percentage of low (high) types increases, the pooling threshold converges to the full information low(high) threshold.

Out-of equilibrium beliefs

Now we turn to the question of which policy interventions can be maintained in a pooling equilibrium. The complication arises from the fact that the equilibrium concept we used so far does not place any restrictions on the beliefs off-the-equilibrium path, which are never reached to verify those beliefs. First, we assume that agents' beliefs are characterized as follows:

- *(Equilibrium path)* If agents observe the pooling level intervention, s_p they play according to respective coordination game, as defined above
- *(Off-the-equilibrium path)* if agents observe any other intervention $s \neq s_p$ they believe that the regime is of low type

This belief system is used very often in the signalling literature as a benchmark. Under this specification, every level of intervention which gives at least as much profit *to both types* what they would get under 'low type' beliefs can be maintained. Since the privately optimal levels for the two types are different, there is no Pareto-dominant pooling equilibrium. [17]. However, any policy intervention $s_p \notin \left[s_p^{L\star}, s_p^{H\star}\right]$ is Pareto-dominated by some intervention level $s_p \in \left[s_p^{L\star}, s_p^{H\star}\right]$.

[17]This is in contrast to, for example, a simple Spence-model, where zero education by both types Pareto-dominates all other pooling equilibria

Which pooling level can be maintained?

The set of pooling equilibria is limited by the fact that in any equilibrium both types should get at least their minimax payoff. Thus any candidate equilibrium s_p is such that

$$\pi^{\star}_{L} \leq \pi^{P}_{L}(s_p)$$
$$\pi^{\star}_{H.L} \leq \pi^{P}_{H}(s_p)$$

These incentive compatibility constraints select a critical value - denoted s_p^{max} as a maximum incentive-compatible signal in a pooling equilibrium.[18] IC(1) above is analytically solved in Appendix. I suspect but it is not proven yet that (2) will never be binding.

Intuitive characterization of pooling equilibrium

The following is an intuitive characterization of the pooling equilibrium

- In any pooling equilibrium, the realized failure threshold for low/high types is still different, with
$$\hat{\theta}^{H}_{2P} < \hat{\theta}^{L}_{2P}$$
- Whenever $p \in (0,1)$, we also have
$$\hat{\theta}^{H}_{2} < \hat{\theta}^{H}_{2P} \qquad \text{and} \qquad \hat{\theta}^{L}_{2P} < \hat{\theta}^{L}_{2}$$
where the respective 'full information' thresholds are calculated with the same hypothetical cash holdings [19] This can be interpreted as some sort of 'risk-sharing' - by pooling with good types, bad types can decrease their failure threshold and survive in more states of the world.
- In the limits $p = 0$ (no bad types) implies $\hat{\theta}^{H}_{2P} = \hat{\theta}^{H}_{2}$ while $p = 1$ (no good types) implies $\hat{\theta}^{L}_{2} = \hat{\theta}^{L}_{2P}$. More generally, the thresholds approach the full-information high(low) threshold if the percentage of high-type agents is high(low).
- More precise information decreases the difference between pooling thresholds. This follows immediately from the analytical expression for $\Delta\theta^{P}_{2}$. This relationship implies that the possible risk-sharing benefits from being in a pooling equilibrium is bigger (for the bad type) when the private information is more precise.
- If the percentage of good types is relatively large, the pooling equilibrium threshold for the good banks is close to the first-best threshold, and the risk-sharing benefit for the bad banks is relatively large and increases with precision. Specifically, a

[18] Precisely, they will select also an s_p^{min} and for the lower boundary the first IC may be binding, but the upper boundary is the interesting one. TBD

[19] i.e. this would be the failure threshold for the given cash holding, if type θ_1 would be known at no cost by all receivers.

pooling on (or close to) the good types first-best optimum intervention is a desirable outcome. However, in precisely those parameter settings the separating equilibrium is wasteful - the distortion from first-best are large in a separating equilibrium precisely if the noise is small, and in addition, for low p a large fraction of types would choose inefficient intervention.

Payoffs under pooling equilibrium. Given the fundamental thresholds we can characterize sender's payoff. For any $i \in \{L, H\}$:

$$\Pi_i^P = \int_{\hat{\theta}_i^P}^{\overline{\theta}_2} k - cs^i \mu(\cdot) d\theta = (k - cs^i)\left(\overline{\theta}_2 - \hat{\theta}_i^P\right)$$

Given the formula for the threshold, this can be written as

$$\Pi_L^P(s) = \Pi_L^{FI}(s) + (k - c \cdot s)\left(\frac{sp_H \Delta\theta_1}{1 + 2\sigma}\right)$$
$$\Pi_H^P(s) = \Pi_H^{FI}(s) - (k - c \cdot s)\left(\frac{sp_L \Delta\theta_1}{1 + 2\sigma}\right)$$

The difference between the two types' profit

$$\Delta\Pi = \frac{2\sigma}{1 + 2\sigma}\left(\theta_1^H - \theta_1^L\right)(k - c \cdot s)s$$
$$\frac{\partial \Delta\Pi}{\partial s} = \frac{2\sigma}{1 + 2\sigma}\left(\theta_1^H - \theta_1^L\right)(k - 2c \cdot s)$$

The profit difference is increasing in s as long as $\frac{k}{2c} > s$ and decreasing thereafter. The total payoff given the prior type distribution is

$$\sum \Pi = (k - c \cdot s)\left(\overline{\theta}_2 - (p_L \hat{\theta}_2^L + p_H \hat{\theta}_2^H)\right)$$

we can rewrite the second term

$$(p_L \hat{\theta}_2^L + p_H \hat{\theta}_2^H) = t - \frac{s(p_L \theta_1^L) + p_H \theta_1^H)}{1 + 2\sigma} - \frac{2\sigma s}{1 + 2\sigma}\left(p_L \theta_1^L + p_H \theta_1^H\right) = t - \left(p_L \theta_1^L + p_H \theta_1^H\right) s$$

Implication: the total payoff to the sender in a pooling equilibrium equals the total profit which would occur with full information, for the given unique intervention level. That is, although agents are unable to distinguish the two types, the sender' ex-ante expected profit is exactly as if they were distinguishable. The only welfare loss (on the sender's side) stems from the fact that in a pooling equilibrium they are unable to set their first-best. This difference can be quantified

$$\Delta\Pi = \Pi(s^\star) - \Pi^s = (k - c \cdot s^\star)(\overline{\theta}_2 + s^\star \theta_1 - t) - (k - c \cdot s)(\overline{\theta}_2 + s\theta_1 - t)$$
$$= k\theta_1 \Delta s - c(\overline{\theta}_2 - t)\Delta s - c\theta_1 (\Delta s)^2$$

4 Signalling with restricted action space

4.1 Regulation

In this section we show how a *minimum policy* regulation can squeeze out the separating equilibrium. Let's denote the minimum policy chosen by regulator as s_p, which means, the action space of the sender is restricted to the interval $s \in [s_p, \infty)$.

Minimum policy changes minimax payoffs for both types. Assume first that minimum policy is high enough to be binding for both types in the sense that it exceeds their minimax strategies, i.e.

$$s_p \geq s^{\star}_{H.L}$$

This single condition is sufficient, since $s^{\star}_{L} < s^{\star}_{H.L}$, so the constraint will always be binding for L-type. Then, we can reformulate IC's for a separating equilibrium as

$$\pi^{eq}(s_p, \theta_1^L) \geq \pi^{off}(s^H, \theta_1^L, \theta_1^H) \qquad (IC_L)$$

$$\pi^{eq}(s_H, \theta_1^H) \geq \pi^{off}(s_p, \theta_1^H, \theta_1^L) \qquad (IC_H)$$

where s^H still denotes the signal a H-type sends in a separating equilibrium. Similarly to the previous (no-regulation) case, we can define $s_L^{cri}(s_p)$ and $s_H^{cri}(s_p)$ as the value of s which solves the two IC's respectively with equality, now regarded as a function of regulatory threshold s_p. Then, the critical regulatory threshold level which guarantees that separating equilibrium does not exist will be determined by the equation:

$$s_L^{cri}(s_p) = s_H^{cri}(s_p) \qquad (13)$$

In Appendix we show that the analytical solution to this equation is:

$$s_p^{cri} = \left(\frac{(1+2\sigma)(\overline{\theta}_2 - t)}{\Delta\theta_1} + \frac{k}{2c}\right) - \sqrt{\left[\frac{(1+2\sigma)(\overline{\theta}_2 - t)}{\Delta\theta_1}\right]^2 + \left[\frac{k}{2c}\right]^2} \qquad (14)$$

Theorem 2 *Let s_p^{cri} denote the critical regulatory minimum as defined under equation 14. As long as the regulatory minimum policy s_p satisfies the condition*

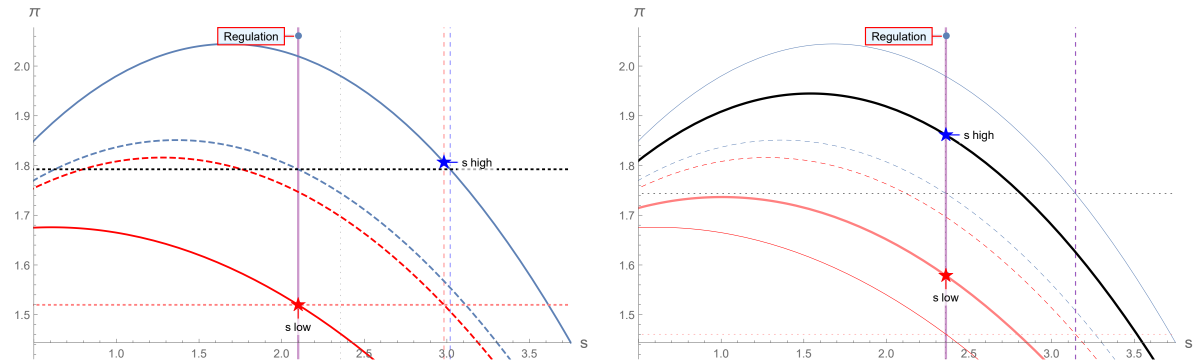
$$s_p \geq s_p^{cri}$$

, separating equilibrium does not exist. The pooling equilibrium in which all types of senders set s_p^{cri} Pareto-dominates from the senders' perspective all other pooling equilibria.

Proof. See Appendix ■

Figure 5a depicts a situation, where despite a quantitative minimum regulation for policy is in place, separating equilibrium still survives, as s_L^{cri} (the policy which is just incentive-compatible for the low type) is lower than s_H^{cri} (just incentive-compatible for the high type). The regulator must increase the minimum policy to at least s_p^{cri}, where $s_L^{cri} = s_H^{cri}$, so separating equilibrium does not exists anymore (figure 5b).

Figure 5: Impact of minimum policy regulation on the equilibrium



(a) Regulation is ineffective: separating equilibrium survives

(b) Regulation is effective: separating equilibrium fails

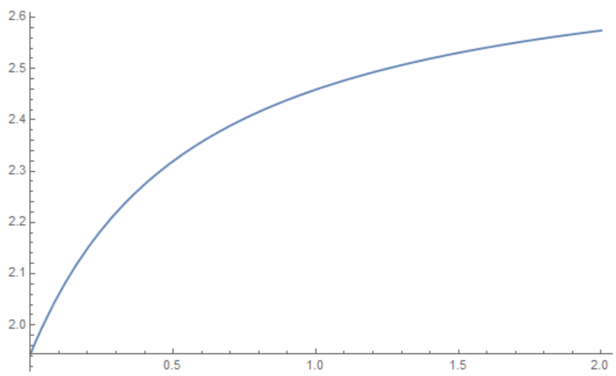
Now we state a result which have interesting implications to the effect of regulation during a crisis situation.

Corollary 2 *The critical regulatory threshold decreases in precision* $1/\sigma$

The intuition behind this result is straightforward. We have seen that with high precision, the first-type inefficiency in a separating equilibrium is very high as high-quality institutions are sending excessively highly signals to distinguish themselves from low-quality institutions. However, this is exactly the situation when it is relatively easy for a regulator to squeeze out the separating equilibrium by setting a relatively low pooling threshold. Since the incentive compatibility constraints for the H-type are already close to binding, a little bit more pressure induced by the regulator can be sufficient for breaking down separation.

Figure 6 illustrates the level of critical regulatory minimum as a function of noise (σ) in private information.

Figure 6: Critical regulation increases in noise in private information



It is possible to interpret this finding in terms of the cyclicality of banking regulation. Noisy private signals represent turbulent economic periods, as increased uncertainty amplifies strategic uncertainty and information asymmetries among creditors. In this case there is relative little to gain from mimicking the other type, so it is easier to maintain separating equilibrium: a regulator must set relatively high minimum policy to be able to squeeze it out. In contrast, in 'normal times' - represented by small noise - there is more temptation to mimic good types, separating levels are highly inefficient, but relatively easy to squeeze them out. This can explain why - observationally - signalling seems to be more prevalent in turbulent times. A regulation which is just sufficient to impose pooling in a normal market environment, might not be sufficient to achieve the same during turbulent times when strategic uncertainty amplifies, and financial institutions engage more and more in costly signalling as the markets shifts towards a crisis period.

4.2 Welfare analysis

In this section we analyse ex-ante expected payoffs for sender and receiver separately, identify sender-optimal and receiver-optimal equilibrium, and draw some conclusions regarding welfare effects of the regulation.

For the ease of exposition let's denote any given equilibrium as

$$Q := \{s_i, s_j, \hat{\theta}_2^i, \hat{\theta}_2^j, \hat{x}^i, \hat{x}^j\}$$

where $\{s_i, s_j\}$ are equilibrium first-stage strategies of a sender of type $\{i, j\}$, $\{\hat{x}_i, \hat{x}_j\}$ are equilibrium strategies of receivers, and $\{\hat{\theta}_2^i, \hat{\theta}_2^j, \}$ describes failure threshold of type i, j. [20]

4.2.1 Sender's ex ante utility

Let π_i^Q (resp. π_j^Q) denote the low (high) type's ex-post[21] *expected* payoff in any equilibrium Q, and by p the probability mass of low-type senders. Then ex-ante expected utility is

$$\mathbb{E}_0 \pi^Q = p\pi_i^Q + (1-p)\pi_j^Q$$

Recall that (irrespectively how the threshold and in turn the probability of survival is calculated) the sender's utility can always be written generally as

$$\pi_t(Q) = \rho(k - cs) = \frac{1}{2\eta}\left(\theta_2^{max} - \hat{\theta}_2(s)\right)(k - cs)$$

where the variables $\hat{\theta}_2$ and s are the ones which are described by equilibrium Q to type $t \in \{i, j\}$. Substituting to the ex-ante formula shows that total ex-ante payoff is a decreasing

[20]If Q denotes a pooling equilibrium, $\{s_i = s_j\}$ and $\{\hat{x}_i = \hat{x}_j\}$

[21]By ex-post we mean after observing private information

function of the average of thresholds in any given equilibrium [22]. The threshold $\hat{\theta}_2(s)$ is always a decreasing function of s (ceteris paribus a sender survives more states with higher s). The expected profit can be characterized by the derivative:

$$\frac{\partial \mathbb{E}\pi}{\partial s} = -k\frac{\partial \overline{\hat{\theta}_2}}{\partial s} - c\theta^{max} + cs\frac{\partial \overline{\hat{\theta}_2}}{\partial s} + c\overline{\hat{\theta}_2} = \frac{\partial \overline{\hat{\theta}_2}}{\partial s}(-k + cs) - c(\theta^{max} - \overline{\hat{\theta}_2}) \tag{15}$$

These observations turn out to be useful to characterize pooling and separating equilibria in the following sections.

Pooling equilibrium

In any pooling equilibrium we have $\overline{\hat{\theta}_2} = t - s\overline{\theta}_1$ so the partial derivative 15 is $-\overline{\theta}_1$ and the total expected payoff as a function of s is

$$\frac{\partial \mathbb{E}\pi}{\partial s} = \overline{\theta}_1(k - cs) - c(\theta^{max} - t + s\overline{\theta}_1)$$

This allows us to calculate the *pooling* equilibrium which maximizes total sender welfare: this is pooling on the value of s which solves

$$\overline{\theta}_1(k - cs) = c(\theta^{max} - t + s\overline{\theta}_1)$$

so the optimal pooling level would be

$$s^\star_{pool} = -\frac{c(\theta^{max} - t) - \overline{\theta}_1 k}{2c\overline{\theta}_1} = \frac{1}{2}\left(\frac{k}{c} - \frac{\theta^{max} - t}{\overline{\theta}_1}\right)$$

It is obvious that $s^\star_i < s^\star_{pool} < s^\star_j$, that is, the best pooling equilibrium is between the first-best intervention levels.

To analyse the effect of a regulatory-imposed pooling, we calculate the equilibrium in which both types set an arbitrary s_p. We know that in pooling the total payoff equals to the weighted total payoff, so we can write

$$\mathbb{E}\pi^P = p\pi^i(s_p) + (1 - p)\pi^j(s_p) = \pi(\overline{\theta}_1, s_p)$$

This can be computed explicitly by substituting to the generic profit function.

$$\mathbb{E}_0\pi^P = \frac{1}{2\eta}\left(\theta_2^{max} - \hat{\theta}_2(\overline{\theta_1}, s_p)\right)(k - cs)$$

The simple formula facilitates easy comparative statics: the ex-ante expected payoff in pooling is *not* a function of noise, and is decreasing in p [23].

[22] As long as $k > cs$ which we always have to impose to ensure that payoffs are nonnegative

[23] The derivative is $\frac{\partial}{\partial p} = -\frac{1}{2\eta}(k - cs_p)s_p\Delta\theta_1 < 0$

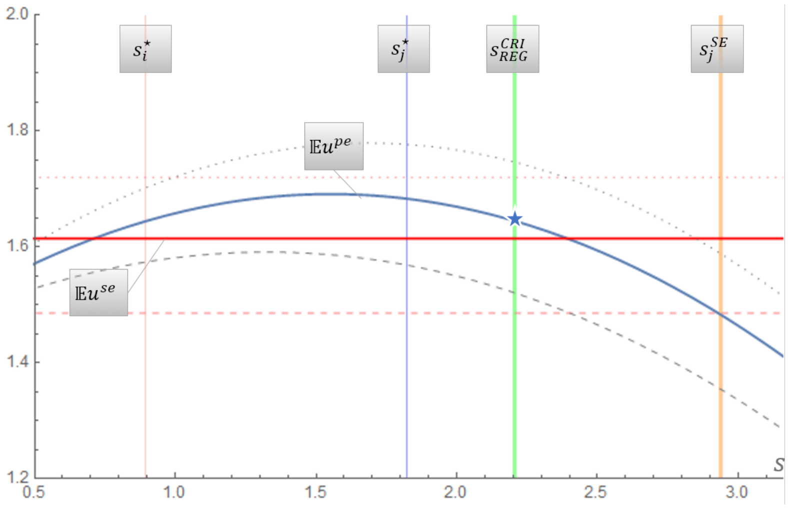


Figure 7: Illustration of senders' welfare

Separating equilibrium

We only focus on the least-costly separating equilibrium, which is well defined by exogenous parameters. In this equilibrium (i) the low-type sends its first-best signal $s_i^\star$ and obtains first-best profit, (ii) the high-type sets the critical value $s_{i.2}^{cri}$ and obtains a profit weakly less than its first-best.

The total payoff to sender is always (weakly) below the first-best:

$$\mathbb{E}_0 U^{SE} = p\pi_i(s_i^\star) + (1-p)\pi_j(s_{i.2}^{cri})$$

The total welfare loss compared to the first-best is concentrated to the high-type:

$$LS^{SE} = p_j(\Pi_j^\star - \Pi_j(s_{i.2}^{cri}))$$

Sender-optimal equilibrium

Putting together, the sender would prefer pooling on s_p over separating if and only if

$$\mathbb{E}_0 U^{PE}(s_p) > \mathbb{E}_0 U^{SE}$$

This can be analytically calculated since we have analytical expressions for all variables.

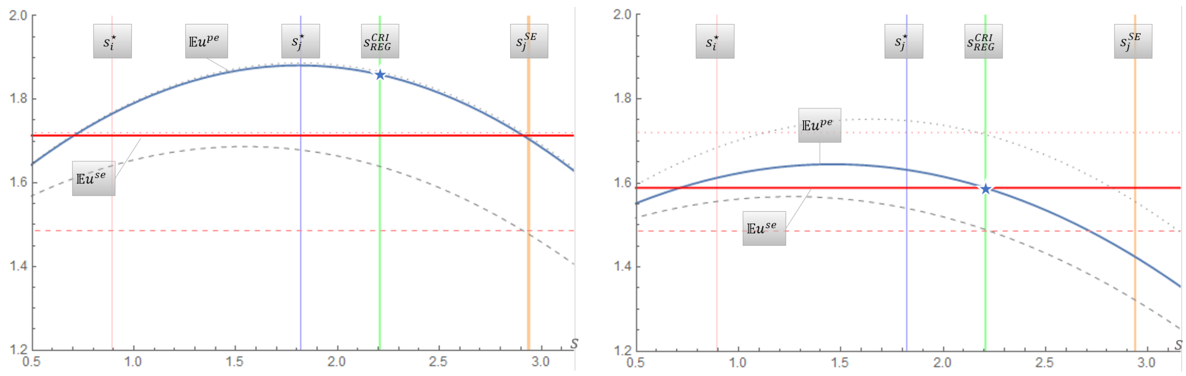
Characterization by numerical analysis

Although analytical analysis is hard to perform due to the complicated formula for separating equilibrium payoffs, the following - intuitive - results are revealed by numerical analysis.

- The ex-ante expected pooling profit is a concave function of s_p (red curve in Figure 7). As long as the 'optimal pooling' exceeds profits from separating equilibrium, there exists a compact interval of interventions at which pooling is preferred from ex-ante perspective to separation. We focus on the upper threshold of this region and introduce notation s_p^{cri}. If it exists, pooling is preferred by sender for $s_p < s_p^{cri}$

- s_p^{cri} strictly decreases in p. In the limit where $p \to 0$ (all types are 'good'), $s_p^{cri} = s_{i.2}^{cri}$, that is, the level of cash set by high type in a separating equilibrium, which means, every pooling below is welfare-improving (Figure 8a). In the other limit, as $p \to 1$, $s_p^{cri} \to s_i^*$ and pooling is not welfare-improving. There exists a critical probability level $p^\star$ such that critical regulation improves welfare if and only if $p < p^\star$ (Figure 8b)

- s_p^{cri} strictly decreases in noise σ. There exists a critical noise σ^{cri} that $\exists s_p^{cri} \Rightarrow \sigma < \sigma^{cri}$



(a) Sender ex-ante payoff as $p \to 0$ (b) Sender ex-ante payoff as $p \to p^\star$

Figure 8: Sender welfare

More intuitively, we can tell the following story: (i) Ex ante, senders *may* (but do not necessarily) prefer pooling on not too large signals. This prevents sending inefficiently high signals in a separating equilibrium. (ii) Lower noise/ higher precision makes high-type-inefficiencies larger. This makes senders to like pooling even more / prefer pooling at even lower levels. In addition, lower noise/higher precision pushes down the minimum threshold which is required to force out separating equilibria. (iii) If there is more good types, the inefficiencies happens more often, so senders prefer pooling even more.

4.2.2 Receivers' ex ante utility

The receivers' payoff is t if she withdraws, 1 if the regime survives and she stays and 0 if the regime fails. Their expected payoff reflects that (i) higher signal has a fundamental stabilizing role, so ceteris paribus they always prefer higher signals (ii) when comparing pooling with separation, the benefits of high-types high signals are weighed against low-type low signal.

Separating equilibrium

In Appendix we derive receivers' ex-ante expected payoff. Because in any separating equilibrium types are perfectly revealed, the ex-ante payoff is the probabilty-weighted average of full-information payoffs using the prior probability distribution given by p.

Lemma 4 *The expected payoff to all receivers facing a sender with known (or correctly deduced) type who sets policy s is*

$$\mathbb{E}_1 u = \frac{1}{2\eta}\left((\theta^{max} - t\theta_{min}) - \sigma t(1-t) - (t - \theta_1 s)(1-t)\right)$$

Proof. See Appendix ■

We are particularly interested in the unique, *least-costly separating equilibrium*, in which the signals set by low(high) type respectively are determined by only exogenous parameters ($s_i^\star$ and $s_{i.2}^{cri}$ respectively).

The ex-ante expected payoff in the least-costly separating equilibrium is

$$\begin{aligned}\mathbb{E}_0 u &= p\mathbb{E}_1 u(s_i^\star) + (1-p)\mathbb{E}_1 u(s_{i.2}^{cri}) \\ &= \frac{1}{2\eta}\left((\theta^{max} - t\theta_{min}) - \sigma t(1-t) - (1-t)\left(t - p\theta_1^i s_i^\star - (1-p)\theta_1^j s_{i.2}^{cri}\right)\right)\end{aligned}$$

Discussion

- Ceteris paribus the noise in agent's signal decreases payoff: a lower precision increases strategic threshold, while keeping the fundamental threshold fixed. That increases the likelihood of making both types of errors (withdrawal if survive, stay if fails).
- Ceteris paribus agents prefer higher signals because of its stabilizing role. In separating equilibrium the high-type sets higher cash which improves payoffs, but the low-type sets low cash which decreases payoffs compared to pooling.
- Precision have another indirect effect with the same direction in a separating equilibrium: higher precision increases the separating level of signal, which is bad for the sender, but good for the receiver -

Pooling equilibrium

Expected payoff in a pooling equilibrium is more complicated because we have to take into account that the two types of senders are facing a mass of agents who follow the same strategy, but they have different failure thresholds. The calculations are however fairly straightforward using the expressions derived under Pooling Equilibrium section. All calculations are included in Appendix)

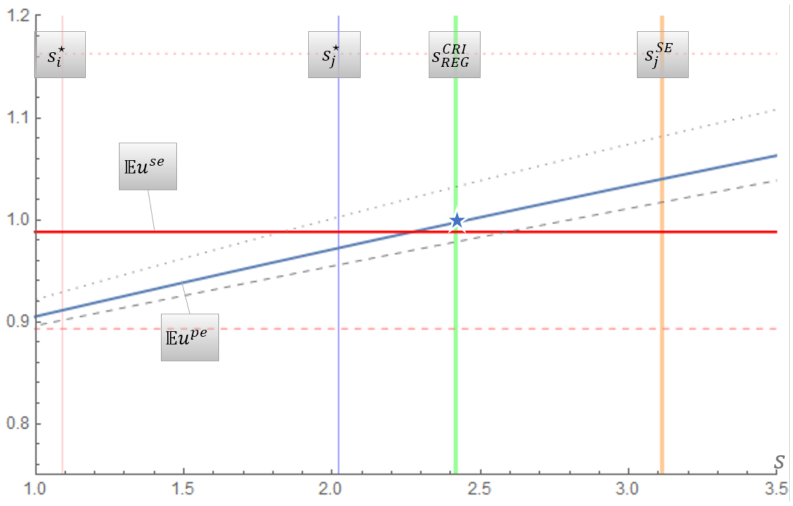


Figure 9: Illustration of senders' welfare

Lemma 5 *In any feasible pooling equilibrium when types pool at signal s, receivers facing with type $\{i, j\}$ respectively obtain the following expected payoff*

$$\mathbb{E}_1 u^i == \frac{1}{2\eta}\left((t^{max} - t\theta_{min}) + (t^2(1+2\sigma) - t\bar{\theta}_1 s - 2\sigma t) + t\sigma - \sigma(t - \frac{sp_j\Delta\theta_1}{1+2\sigma})^2 - \hat{\theta}_P^i\right)$$

and

$$\mathbb{E}_1 u^j = \frac{1}{2\eta}\left((t^{max} - t\theta_{min}) + (t^2(1+2\sigma) - t\bar{\theta}_1 s - 2\sigma t) + t\sigma - \sigma(t + \frac{sp_i\Delta\theta_1}{1+2\sigma})^2 - \hat{\theta}_P^j\right)$$

The ex-ante expected payoff is therefore

$$\mathbb{E}_0 u = \frac{1}{2\eta}\left((t^{max} - t\theta_{min}) + t^2(1+2\sigma) + \bar{\theta}_1 s(1-t) - \sigma t - \sigma p_i\left(t - \frac{sp_j\Delta\theta_1}{1+2\sigma}\right)^2 - \sigma p_j\left(t + \frac{sp_i\Delta\theta_1}{1+2\sigma}\right)^2 - t\right)$$

It is possible to analytically determine the region where a pooling equilibrium is preferred by receivers from ex-ante point of view, but the calculations are tedious.

Discussion

- Separating and pooling profits and ex-ante type distribution:
  - when $p = 1$ (all types are low-types), the SE-payoff equals to the PE-payoff if s^p is set to $s_i^\star$. Every higher s_p improves payoff to the receivers.
  - when $p = 0$ (all types are high-types), the SE-payoff equals to the pooling payoff if pooling were set to the separating level cash. In this situation imposing pooling at any level $s^p < s_i^{cri}$ reduces payoff to receivers (but leaves more profit to the banks)

- In every point between those extremes, there exist a pooling level s_p^0 which is just sufficient to make sure that $\mathbb{E}u^{SE} < \mathbb{E}u^{PE}$. This value is decreasing in p (with more low-type sender, a lower pooling level is sufficient to increase utility of receivers).
- A remaining question to establish is how this level relates to the level of s_p, denoted by s_p^{cri} which is sufficient to squeeze out separating equilibrium completely.

• Separating and pooling profits and noise precision

  - Receivers' payoff is increasing in precision in all equilibria. This is due to the decrease in both types of errors what creditors make due to the effective strategy.
  - Higher precision means higher separating signal for the high type, as established in 'separating equilibrium' chapter. That means, a higher s^p is required to 'compensate' creditors. Higher precision pushes s_p^0 upwards.

5 Empirical analysis

We illustrate our model with two simple empirical analysis. The first example builds on the recent introduction of Liquidity Coverage Ratio (LCR) in the US. In essence, the LCR places a quantitative lower bound on the amount of liquid assets which must be held by financial institutions at all time. In the analysis we exploit the fact that the rule only applies to 'large, internationally active' bank holding companies and show that - consistently with the idea of squeezing out separating equilibrium - the introduction of LCR regulation was followed by a larger decrease of volatility of cash ratios for holding companies which were subject to the newly introduced regulation, compared to those below the threshold of qualifying for regulation.

Our second example investigates changes in equity ratios around the introduction of the first generation of Basel capital regulations in 1988. Interestingly, the data shows a significant increase of the number of banks with high capital ratios after the introduction of the new regulatory regime. According to our theory, this is consistent with a quantitative regulatory requirement set too low, and therefore being unable to squeeze out separating equilibrium.

5.1 Liquidity Coverage Ratio

As part of its regulatory reform package in response to the financial crisis known as Basel III, the Basel Committee of Banking Supervision has put forward a series of measures concerning the liquidity risk framework of financial institutions. The agenda consists of two key elements: the Liquidity Coverage Ratio (LCR) requires banks to hold adequate amount of highly liquid assets to cover outflows in a crisis scenario over a 30 days period, while the Net Stable Funding Ratio (NSFR) supplements this measure by ensuring a sustainable asset-liability maturity structure over a longer time horizon. In our analysis we focus on the former measure.

The Basel Committee announced final version of LCR in January, 2013[24] and adopted a gradual approach for implementation, with the full version in effect from January 2019. In November 2013, US authorities proposed an LCR regulation largely consistent with the guidelines set forth by the Basel Committee. The final rule has been adopted in September 2014, being in effect from January 2015, with a much shorter transition period than the Basel III proposal *"to preserve the strong liquidity positions many U.S. banking organizations have achieved since the recent financial crisis"*.

For our analysis, we have constructed a 'treatment' sample consisting of internationally active bank holding companies (BHC's) with Total Assets more than \$50bn, and we look at changes of cash ratios (defined as Cash / Total Assets) from 2011 to 2016, based

[24]Basel III, B.C.B.S. "The Liquidity Coverage Ratio and liquidity risk monitoring tools." Bank for International Settlements (2013).

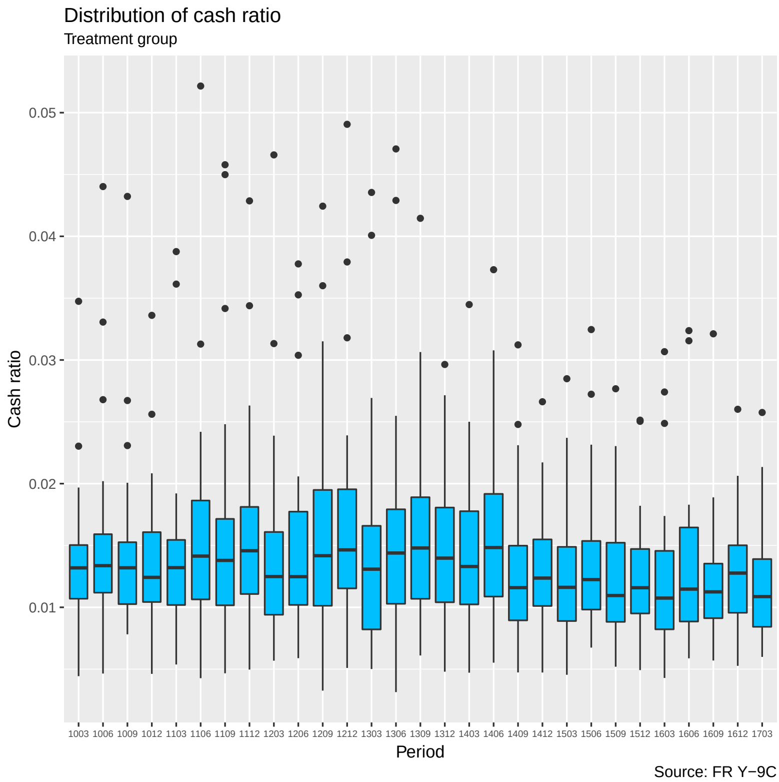


Figure 10: Timeline

on FR-Y-9C filings. Our control sample consists of BHC's with Total Assets between $10bn and $40bn, which consistently reported throughout the whole sample period. Table 1 summarizes treatment and control sample.

	treatment	control
nrBanks	32	73
avgCashRat	0.0142	0.0143
stDevCashRat	0.0067	0.0082
totalAsset	441 765 936	16 799 271

Table 1: Overview of treatment and control samples

Findings

We find that both mean and standard deviation of cash ratios decreased during the period for both the treatment and the control sample. However, following the announcement of Basel III LCR, and especially around the introduction of the follow-up US regulation, the standard deviation of cash ratios in the treatment sample started to decrease significantly more sharply than that of the control sample (see Figure 11). This was mainly driven by the disappearance of larger values (i.e. BHC's with too high cash ratios, see Figure 10), which, in the context of our model, can be interpreted as an elimination of separating equilibrium. We plot the six BHC's with the biggest cash drops on average before/after the event date in Figure 12a. To further emphasize this finding, in Figure 12b

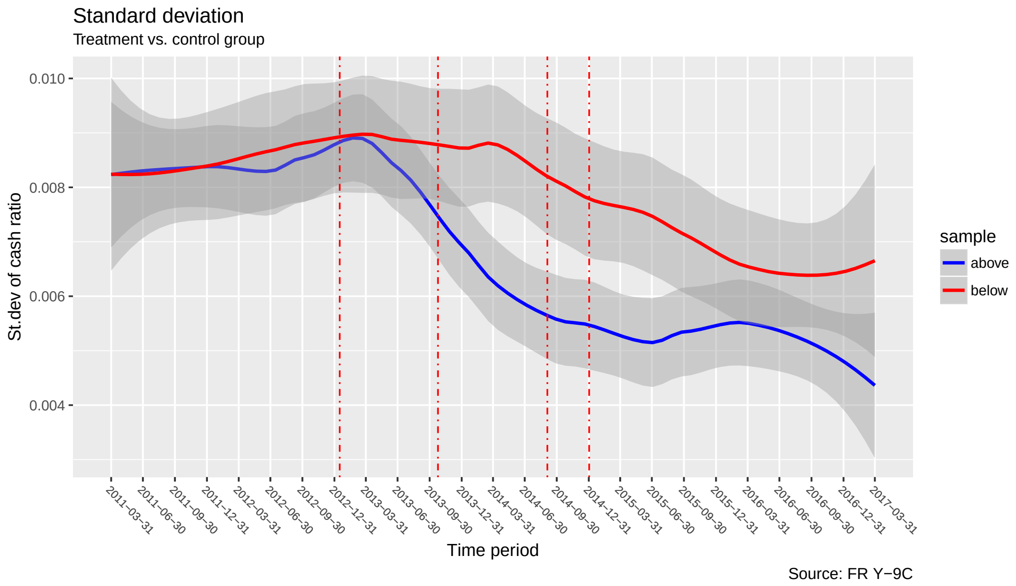
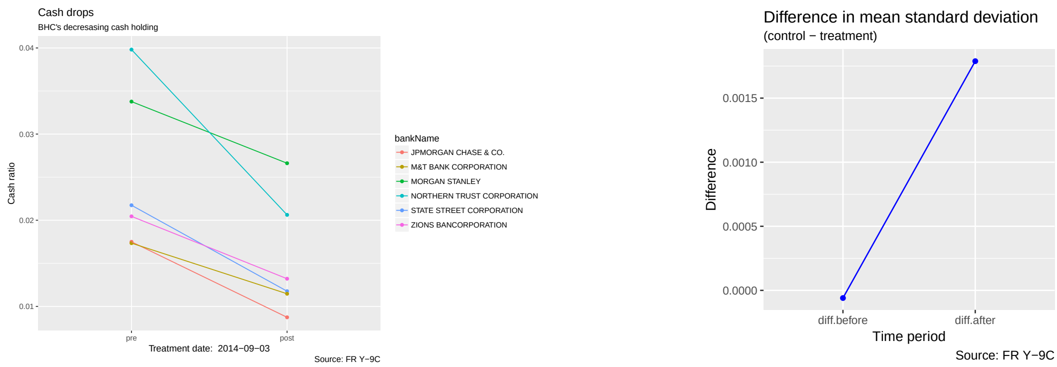


Figure 11: Timeline

we plot the difference of average standard deviation for the sample and treatment group before and after the date of announcement of US regulation. The increasing difference justifies the larger clustering of observation in the treatment sample relative to control as an effect of introducing liquidity regulation.

Finally, we perform a Kolmogorov-Smirnov test to mechanically compare the distributions of treatment and control samples. Before the announcement of LCR regulation, we cannot reject the null-hypothesis that the two distributions are the same (p-value: 0.79), while it can be rejected after the event (p-value: 0.01).



(a) (b)

Figure 12: Impact of minimum policy regulation on equilibrium

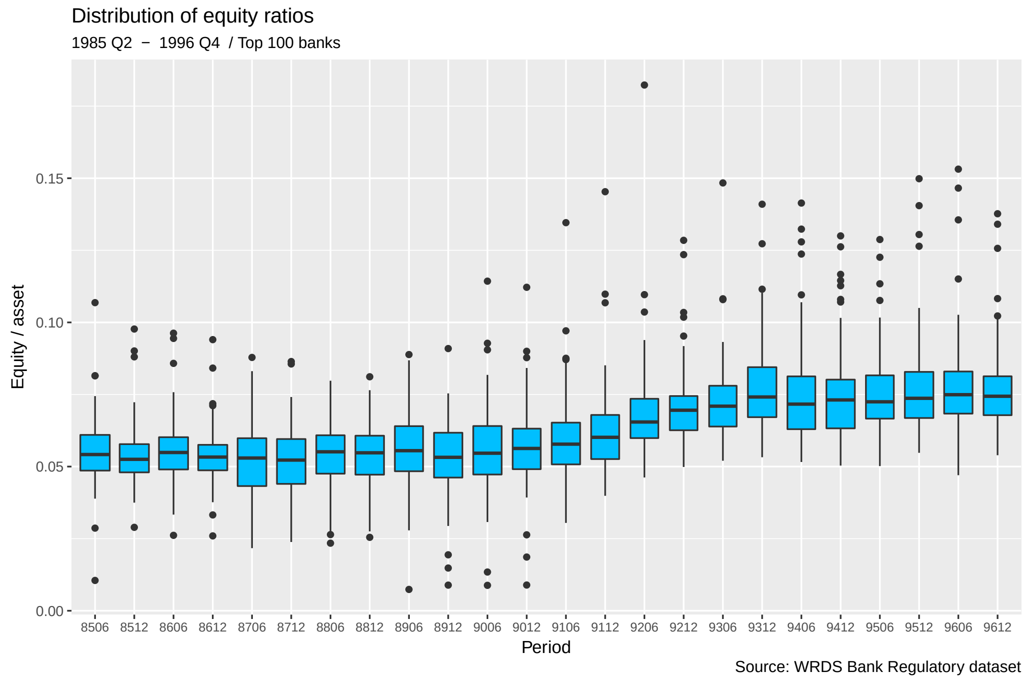


Figure 13: Distribution of equity ratios

5.2 Basel I Capital Regulation

Our second empirical analysis focuses on the introduction of minimum capital regulations under Basel I. The Basel Committee of Banking Supervision (BCBS) published the requirements in 1988 and required banks to maintain a minimum ratio of capital over total risk-weighted asset. Although we do not attempt reconstruct the nominator (weighted sum of various elements of banks' capital) or the denominator (risk-weighted assets), we believe that for our purposes the plain equity ratio defined as

$$er := \frac{\text{Total Equity}}{\text{Total Assets}}$$

will suffice. Figure 13 plots the distribution of equity ratios of the top 100 (by Total Asset) US-regulated bank in the period 1985-1996. Even without any formal statistical analysis, a significant structural change is recognizable around the introduction of Basel I (although announced in 1988, the capital requirements were binding from 1992). First, the bottom-end of the distributions notably shifted upwards, consistently with the regulatory intention behind the new set of rules.[25] Another visible characteristics however, which was certainly not an explicit regulatory aim, an increase in variance, and especially an increase of the number of 'outliers', i.e. institutions maintaining significantly larger equity ratios.

[25]Note that this notable upward shift is missing from the time series of distributions of cash ratios. This is due to the fact that cash holdings have significantly increased during the financial crisis, but already were on a downward trajectory, and one of the reasons of the swift introduction of LCR in the US was to prevent the elevated levels to fall below pre-crisis levels.

In the context of our model, this effect is consistent with a quantitative regulatory minimum which is not sufficiently large to squeeze out separating equilibrium. Indeed, as we have shown in Section 4.1, an insufficiently high regulatory minimum can preserve the incentives for signalling and even increase the minimum signal which is required to maintain separation.

6 Conclusions

In this paper we proposed a model which combines signalling and global games to understand the informational impact of risk management measures and the possible effect of regulation in banking. We established two main results: (i) absent regulation, banks have incentives to signal their quality and may engage in 'excess risk management', which is inefficient; (ii) a financial regulation can squeeze out inefficient separation and improve welfare by introducing a quantitative *minimum* of the given risk measure. Our results provide a novel perspective on understanding some consequences of financial regulation.

The model has testable empirical predictions: for example, in the context of liquidity holdings, in the absence of quantitative regulation in place, we expect high dispersion of liquidity ratios, which is consistently found in bank as well as mutual fund databases. Introducing liquidity regulation neutralizes this incentives, so we expect clustering of observations around the requirement. We find that changes in cash ratios following the recent introduction of LCR ratios in the US are consistent with this hypothesis. Changes in equity ratios, however, around the introduction of Basel I regulatory capital regime are rather consistent with an insufficiently large regulatory minimum, unable to squeeze out separation.

References

[1] George-Marios Angeletos and Alessandro Pavan. "Selection-free predictions in global games with endogenous information and multiple equilibria". In: *Theoretical Economics* 8.3 (2013), pp. 883–938.

[2] George-Marios Angeletos, Christian Hellwig, and Alessandro Pavan. "Signaling in a global game: Coordination and policy traps". In: *Journal of Political economy* 114.3 (2006), pp. 452–484.

[3] William H Beaver and Ellen E Engel. "Discretionary behavior with respect to allowances for loan losses and the behavior of security prices". In: *Journal of Accounting and Economics* 22.1 (1996), pp. 177–206.

[4] Richard Brealey, Hayne E Leland, and David H Pyle. "Informational asymmetries, financial structure, and financial intermediation". In: *The journal of Finance* 32.2 (1977), pp. 371–387.

[5] Hans Carlsson and Eric Van Damme. "Global games and equilibrium selection". In: *Econometrica: Journal of the Econometric Society* (1993), pp. 989–1018.

[6] Douglas W Diamond and Philip H Dybvig. "Bank runs, deposit insurance, and liquidity". In: *The journal of political economy* (1983), pp. 401–419.

[7] Douglas W Diamond and Anil K Kashyap. "Liquidity Requirements, Liquidity Choice, and Financial Stability". In: *Handbook of Macroeconomics* 2 (2016), pp. 2263–2303.

[8] Douglas W Diamond and Raghuram G Rajan. "A theory of bank capital". In: *The Journal of Finance* 55.6 (2000), pp. 2431–2465.

[9] Itay Goldstein and Chong Huang. "Bayesian persuasion in coordination games". In: *The American Economic Review* 106.5 (2016), pp. 592–596.

[10] Itay Goldstein and Ady Pauzner. "Demand–deposit contracts and the probability of bank runs". In: *the Journal of Finance* 60.3 (2005), pp. 1293–1327.

[11] Theoharry Grammatikos and Anthony Saunders. "Additions to bank loan-loss reserves: Good news or bad news?" In: *Journal of Monetary Economics* 25.2 (1990), pp. 289–304.

[12] Paul A Griffin and Samoa JR Wallach. "Latin American lending by major US banks: The effects of disclosures about nonaccrual loans and loan loss provisions". In: *Accounting Review* (1991), pp. 830–846.

[13] Milton Harris and Artur Raviv. "The theory of capital structure". In: *the Journal of Finance* 46.1 (1991), pp. 297–355.

[14] Zhiguo He. "The sale of multiple assets with private information". In: *The Review of Financial Studies* 22.11 (2009), pp. 4787–4820.

[15] Joseph P Hughes and Loretta J Mester. "Bank capitalization and cost: Evidence of scale economies in risk management and signaling". In: *The review of Economics and Statistics* 80.2 (1998), pp. 314–325.

[16] Nicolas Inostroza and Alessandro Pavan. "Persuasion in Global Games with Application to Stress Testing". In: *Economist* (2017).

[17] Frédéric Malherbe. "Self-Fulfilling Liquidity Dry-Ups". In: *The Journal of Finance* 69.2 (2014), pp. 947–970.

[18] Stephen Morris and Hyun Song Shin. "Unique equilibrium in a model of self-fulfilling currency attacks". In: *American Economic Review* (1998), pp. 587–597.

[19] Antonio Nicolo and Loriana Pelizzon. "Credit derivatives, capital requirements and opaque OTC markets". In: *Journal of Financial Intermediation* 17.4 (2008), pp. 444–463.

[20] Stephen A Ross. "The determination of financial structure: the incentive-signalling approach". In: *The bell journal of economics* (1977), pp. 23–40.

[21] Myron S Scholes, G Peter Wilson, and Mark A Wolfson. "Tax planning, regulatory capital planning, and financial reporting strategy for commercial banks". In: *The Review of Financial Studies* 3.4 (1990), pp. 625–650.

Appendix A Appendix A - Proofs

Appendix A.1 Proof of Lemma 1

The unique, monotone threshold equilibrium of the global game with known θ_1 is completely characterized by pair of thresholds $\{\hat{\theta}_2, \hat{x}\}$ which jointly solve two equations: (i) an agent who receives signal $\hat{x}$ is just indifferent between attack and not attack, (ii) the regime just fails at $\hat{\theta}_2$. We formalize the two equations and solve for the unique equilibrium.

Equation 1 *(agents' indifference condition)* An agent is indifferent between attack and not attack if the expected payoff difference is zero:

$$Pr[\text{Failure}](-t) + Pr[\text{Survive}](1-t) = 0$$

This simplifies to

$$Pr[\text{Survive}] = t$$

An agent's posterior probability of survival can be written generally as

$$Pr[\theta_2 > \hat{\theta}_2 | x_i] = 1 - Pr[\theta_2 < \hat{\theta}_2 | x_i] = Pr[\epsilon_i < x_i - \hat{\theta}_2]$$

Substituting the CDF of a uniform distribution

$$Pr[\epsilon_i < x_i - \hat{\theta}_2] = \begin{cases} 0 & \text{if } x_i < \hat{\theta}_2 - \sigma \\ \frac{x_i - \hat{\theta}_2 + \sigma}{2\sigma} & \text{if } \hat{\theta}_2 - \sigma \leq \hat{\theta}_2 + \sigma \\ 1 & \text{if } x_i > \hat{\theta}_2 + \sigma \end{cases}$$

The indifference condition must be true for the agent who happens to receive the threshold signal $\hat{x}$, and it only can be true in the intermediate region, so the indifference condition is

$$Pr[\text{Survive}] = \frac{\hat{x} - \hat{\theta}_2 + \epsilon}{2\epsilon} = t \qquad \Leftrightarrow \qquad \hat{x} = \hat{\theta}_2 + 2\sigma t - \sigma$$

Equation 2 *(regime failure condition)*. By definition, the bank must just fail at $\hat{\theta}_2$, whenever creditors follow threshold strategy $\hat{x}$. The mass of agents who attack the regime given any realization θ_2 is

$$\alpha(\hat{x}, \theta_2) = Pr(x_i < \hat{x} | \theta_2) = \begin{cases} 1 & \text{if } \theta_2 < \hat{x} - \sigma \\ \frac{\hat{x} - \theta_2 + \sigma}{2\sigma} & \text{if } \hat{x} - \sigma \leq \theta_2 \leq \hat{x} + \sigma \\ 0 & \text{if } \hat{x} + \sigma < \theta_2 \end{cases}$$

In the intermediate range this is exactly the same expression as the indifference condition. Combining the two gives

$$\alpha(\hat{x}, \hat{\theta}_2) = \frac{\hat{x} - \hat{\theta}_2 + \epsilon}{2\epsilon} = t$$

The fundamental threshold $\hat{\theta}_2$ solves the failure condition

$$\mathcal{R} = \theta_1 s + \theta_2 - \alpha = 0$$

After substituting $\alpha(\hat{x}, \hat{\theta}_2) = t$ we obtain the *fundamental threshold* $\hat{\theta}_2$, and substituting back to the indifference condition (Equation 1) gives the *strategic threshold* $\hat{x}$.

$$\hat{\theta}_2 = t - \theta_1 s$$
$$\hat{x} = t - \theta_1 s + 2\sigma t - \sigma$$

Uniqueness of equilibrium.

First, we establish that $\forall s \in \mathcal{S}$, $\exists \theta_2(s)$ such that $\mathcal{R}(\theta_1, \theta_2(s), 0) < 0$. In this lower dominance region bank defaults even with no runs ($\alpha = 0$). Denote this critical value by H_0.

$$H_0(s) = -\theta_1 s$$

Under the most optimistic beliefs, a creditor whose posterior belief places positive weight on $\theta_2 < H_0(s)$ believes that bank fails if and only if $\theta_2 < H_0(s)$. Under these beliefs there exists a critical signal, denoted by $h_0(s)$ such that all creditors with signal $x_i < h_0(s)$ find it dominant to run on the bank. The critical signal $h_o(s)$ is determined by the indifference condition of that creditor:

$$Pr[Survive|H_o(s)] = t$$

where

$$Pr[Survive] = Pr[\theta_2 > H_0(s)|h_0(s)] = \frac{h_0(s) - H_0(s) + \sigma}{2\sigma}$$

therefore

$$h_0(s) = -\theta_1 s + 2\sigma t - \sigma$$

We established that it is always strictly dominant to run whenever $x_i < h_0(s)$. For $n \geq 1$ define $H_n(s)$ as the value of θ_2 which solves $\mathcal{R} = 0$ given that *only* creditors with $x_i < h_{n-1}(s)$ run on the bank. The mass of attack is

$$\alpha_n = \frac{h_{n-1}(s) - H_n(s) + \sigma}{2\sigma}$$

which is substituted to function $\mathcal{R}$

$$\theta_2 + \theta_1 s - \frac{h_{n-1}(s) - \theta_2 + \sigma}{2\sigma}$$

leading to

$$H_n(s)\left(1+\frac{1}{2\sigma}\right)=\frac{1}{2\sigma}\left(h_{n-1}(s)+\sigma\right)-\theta_1 s$$
$$H_n(s)\left(\frac{2\sigma+1}{2\sigma}\right)=\frac{1}{2\sigma}\left(h_{n-1}(s)+\sigma\right)-\theta_1 s$$
$$H_n(s)=\frac{1}{2\sigma+1}\left(h_{n-1}(s)+\sigma-2\sigma\theta_1 s\right)$$

Similarly, for $n>1$ let's define $h_n(s)$ as the value of signal x where it is strictly dominant to run, given it is believed that the bank fails if and only if $\theta_2 < H_n(s)$

$$\frac{h_n(s)-H_n(s)+\sigma}{2\sigma}=t$$

$$h_n(s)=H_n(s)+2\sigma t-\sigma$$

Iterate the threshold H by one and substitute the expression for $h_n(s)$

$$H_{n+1}(s)=\frac{1}{2\sigma+1}\left(h_n(s)+\sigma-2\sigma\theta_1 s\right)$$

$$H_{n+1}(s)=\frac{1}{2\sigma+1}(H_n(s)+2\sigma t-\sigma+\sigma-2\sigma\theta_1 s)$$

The series is monotonically increasing as long as $H_n(s) < t-\theta_1 s$. However, with $n\to\infty$ the steady state is

$$H(s)(1-\frac{1}{2\sigma+1})=\frac{1}{2\sigma+1}\left(2\sigma t-2\sigma\theta_1 s\right)$$

$$H(s)(\frac{2\sigma}{2\sigma+1})=\frac{1}{2\sigma+1}\left(2\sigma t-2\sigma\theta_1 s\right)$$

$$H(s)=\frac{1}{2\sigma}\left(2\sigma t-\theta_1 s\right)=t-\theta_1 s=\hat{x}$$

This proves that even with the most optimistic beliefs, by iterative elimination of strictly dominant strategies, all rationalizable action profile must be RUN *if* $x_i < \hat{x}$

—

Analogously, it is possible to construct iterative deletion of strictly dominated strategies from above: this will show that *even with the most pessimistic beliefs*, that means, RUN always whenever it is not strictly dominant to WAIT, it is never rationalizable to RUN if $x_i > \hat{x}$. The two parts together implies that the *unique, rationalizable action profile* for creditors is RUN if and only if $x_i < \hat{x}$.

This concludes the proof.

Appendix A.2 Proof of Lemma 2

Calculate the first derivatives with respect to θ_1 for both optimal signal and optimal profit. For the optimal signal:

$$\frac{\partial s^\star}{\partial \theta_1} = \frac{1}{2}\frac{\theta_2^{max} - t}{\theta_1^2}$$

This is always strictly positive due to the natural parametric assumptions $\theta_1^{max} > 1 > t$. For the optimal profit:

$$\frac{\partial \Pi^\star}{\partial \theta} = \frac{k^2\theta_1^2 - [c(t - \theta^{max})]^2}{4c\theta_1^2} = \frac{1}{4}\left(\frac{k^2}{c} - \frac{c(t - \theta_2^{max})^2}{\theta_1^2}\right)$$

the derivative is strictly positive whenever the following condition holds.

$$\frac{k}{c} \geq \frac{\theta_2^{max} - t}{\theta_1}$$

Note that the condition implies $s^\star \geq 0$.

Appendix A.3 Proof of Theorem 1

Recall the two IC's

$$\pi_i(s_i^\star) \geq \pi_{i.j}(s^{SE}) \qquad (IC_i)$$

$$\pi_{j.i}(s_{j.i}^\star) \leq \pi_j(s^{SE}) \qquad (IC_j)$$

where the functions $\pi_i(\cdot), \pi_j(\cdot)$ and $s_i^\star$ are defined in Lemma 1, while $\pi_{i.j}(\cdot), \pi_{j.i}(\cdot)$ and $s_{j.i}^\star$ in equations 8 and 9 in the main text. Using the notation ρ we can rewrite IC_i as

$$\rho_i(s_i^\star)(k - cs_i^\star) - \rho_{i.j}(s)(k - cs) \geq 0$$
$$\frac{1}{2\eta}\left(\theta_2^{max} - \hat{\theta}_2^i(s_i^\star)\right)(k - cs_i^\star) - \frac{1}{2\eta}\left(\theta_2^{max} - \hat{\theta}_2^{i.j}(s)\right)(k - cs) \geq 0$$
$$(k - cs)\left(\hat{\theta}_2^{i.j}(s) - \hat{\theta}_2^i(s_i^\star)\right) + \left(\theta_2^{max} - \hat{\theta}_2^i(s_i^\star)\right)(cs - cs_i^\star) \geq 0$$

while IC_j as

$$\rho_j(s)(k - cs) - \rho_{j.i}(s_{j.i}^\star)(k - cs_{j.i}^\star) \geq 0$$
$$\frac{1}{2\eta}\left(\theta_2^{max} - \hat{\theta}_2^j(s)\right)(k - cs) - \frac{1}{2\eta}\left(\theta_2^{max} - \hat{\theta}_2^{j.i}(s_{j.i}^\star)\right)(k - cs_{j.i}^\star) \geq 0$$
$$(k - cs)\left(\hat{\theta}_2^j(s) - \hat{\theta}_2^{j.i}s_{j.i}^\star\right) + \left(\theta_2^{max} - \hat{\theta}_2^{j.i}s_{j.i}^\star\right)(cs - cs_{j.i}^\star) \geq 0$$

After substituting the known algebraic formulas for thresholds and optimal interventions, we can solve for the critical value of signal which solves the two IC's with equality.

$$s_{i.2}^{cri} = \frac{1}{2}\left(\frac{k}{c} - \frac{(1+2\sigma)(\theta_2^{max}-t)}{2\sigma\theta_1^i+\theta_1^j}\right) + \frac{1}{2c\sqrt{\theta_1^i}}\sqrt{\frac{(1+2\sigma)}{(2\sigma\theta_1^i+\theta_1^j)}}\sqrt{\Delta\theta_1\left(\frac{k^2\theta_1^i}{1+2\sigma} - \frac{c^2(\theta_2^{max}-t)^2}{2\sigma\theta_1^i+\theta_1^j}\right)}$$

$$s_{j.2}^{cri} = \frac{1}{2}\left(\frac{k}{c} - \frac{\theta_2^{max}-t}{\theta_1^j}\right) + \frac{1}{2c\theta_1^j}\sqrt{\Delta\theta_2\left(\frac{k^2\theta_1^j}{1+2\sigma} - \frac{c^2(t-\theta_2^{max})^2}{\theta_1^i+2\sigma\theta_1^j}\right)}$$

Both thresholds are decreasing functions of the agents' noise σ. Start with the limit of large noise. We calculate the limits where $\sigma \to \infty$.

$$\lim_{\sigma\to\infty} s_{i.2}^{cri} = \frac{1}{2}\left(\frac{k}{c} - \frac{(1+2\sigma)(\theta_2^{max}-t)}{2\sigma\theta_1^i+\theta_1^j}\right) + 0 = s_i^\star$$

$$\lim_{\epsilon\to\infty} s_{j.2}^{cri} = \frac{1}{2}\left(\frac{k}{c} - \frac{\theta_2^{max}-t}{\theta_1^j}\right) = s_j^\star$$

Since $s_{i.2}^{cri}$ continuously and monotonically approaches $s_i^\star$ as $\sigma \to 0$, and $s_i^\star < s_j^\star$ by Lemma 2, by the intermediate value theorem there exists a unique value of σ, denoted by $\overline{\sigma}$, such that

$$s_{i.2}^{cri}(\overline{\sigma}) = s_j^\star$$

Whenever $\sigma > \underline{\sigma}$, the pair $\{s_i^\star; s_j^\star\}$ is incentive-compatible for the low (i) and high (j) types as well, and the first-best can be maintained as a separating equilibrium.

Now we turn to the analysis of the case of small noise, as $\sigma \to 0$. First, calculate the limit of critical signals as $\sigma \to 0$. After some algebraic manipulations it is possible to show that the critical incentive-compatible signals converge to the same expression:

$$\lim_{\epsilon\to 0} s_{i.2}^{cri} = \lim_{\epsilon\to 0} s_{j.2}^{cri} = \frac{1}{2}\left(\frac{k}{c} - \frac{\theta_2^{max}-t}{\theta_1^j}\right) + \frac{1}{2c\theta_1^j}\sqrt{\Delta\theta_1\left(k^2\theta_1^j - \frac{c^2(\theta_2^{max}-t)^2}{\theta_1^i}\right)}$$

A 'constructive' approach to prove the statement of the theorem by solving the equation $s_i^{cri} = s_j^{cri}$ is not possible due to the analytical complexity of the non/limiting case. Instead, we prove the theorem using the following steps, which are analytically easier to calculate:

1. Calculate the partial derivatives $\frac{\partial s_i^{cri}}{\partial\sigma}$ and $\frac{\partial s_j^{cri}}{\partial\sigma}$

2. Consider the value of the derivatives at $\sigma = 0$. If $(s_i^c ri)'(0) < (s_j^c ri)'(0)$ then $s_j^c ri$ approaches the limit faster, meaning that for sufficiently small σ, we must have $s_i^c ri > s_j^c ri$.

3. Analytically solve the equation $(s_i^c ri)'(0) = (s_j^c ri)'(0)$ for c. This gives a critical cost level $\hat{c}$ such that $s_i^c ri > s_j^c ri$ for sufficiently small σ

We omit the detailed calculations to save space. The solution for $\hat{c}$ is

$$\hat{c} = \sqrt{\frac{\Delta\theta_1(k\theta_1^i)^2}{\left(3\theta_1^i+\theta_1^j\right)(\theta_2^{max}-t)}} \tag{A.16}$$

Appendix A.4 Proof of Lemma 3

We prove a more general version of Lemma 3, from which the version in the main text will be trivial. In particular, we prove the characterization of pooling equilibrium with an arbitrary N senders.

Suppose the number of senders is $N \geq 2$, each with types θ_1^n. Without loss of generality we can determine the indexing of banks such that $n_i < n_j \Leftrightarrow \theta_1^{n_i} < \theta_i^{n_j}$. Let the prior distribution of types be $Pr[\theta_1 = \theta_1^n] = p_n$. It is useful to interpret $N = 1, 2...$ as *quality classes* and the probability p_n representing the mass of institutions belonging to this quality class.

Suppose that a closed subset of institutions[26] $\mathcal{N} := \underline{n} < n < \overline{n}$ are pooling on the same risk-management signal s_p. Let us define the conditional distribution of banks belonging to $\mathcal{N}$ as $\tilde{\mathcal{P}} := \{\tilde{p}_n\}_{\underline{n}}^{\overline{n}}$, it is straightforward that

$$\tilde{p}_n = \frac{p_n}{\sum_{n \in \mathcal{N}} p_n}$$

Conditional on observing s_p, $\tilde{p}_n$ represents receiver's posterior probability of the event that the sender is of type n. We define the (conditional) average type as

$$\overline{\theta}_1^n := \sum_{n \in \mathcal{N}} \tilde{p}_n \theta_1^n$$

Equation 1: we start with the creditors' indifference condition. A creditor is indifferent between actions WAIT and RUN if

$$Pr[Failure] \cdot \underbrace{(0 - t)}_{\text{p/o wait-run}} + Pr[Survive] \cdot \underbrace{(1 - t)}_{\text{p/o wait-run}} = 0$$

Because $Pr(Failure) = 1 - Pr(Survive)$, we can rewrite this equation as

$$Pr[Survive] = t$$

Using creditors' posterior probability, we can write

$$Pr[Survive] = \sum_{\underline{n}}^{\overline{n}} \tilde{p}_n \Phi \left(\frac{\hat{x} - \hat{\theta}_1^n}{\sigma} \right)$$

After substitution

$$\sum_{\underline{n}}^{\overline{n}} \tilde{p}_n \frac{\hat{x} - \hat{\theta}_1^n + \sigma}{2\sigma} = t$$

[26] actually, it doesn't have to be a closed subset, it can be any subset so far, but I believe this is important later when showing existence of pooling equilibria

this can be rewritten as

$$\sum \tilde{p}_n \hat{x} - \sum \tilde{p}_n \hat{\theta}_1^n + \sum \tilde{p}_n \sigma = 2\sigma t$$
$$\hat{x} - \overline{\hat{\theta}_1^n} + \sigma = 2\sigma t$$
$$\hat{x} = 2\sigma t - \sigma + \overline{\hat{\theta}_1^n}$$

Equation 2: Given strategic threshold $\hat{x}$, the fundamental threshold solves $\mathcal{R} = \theta_1 s + \theta_2 - \alpha = 0$. After substituting α and rearranging the equation, we have for each $n \in \mathcal{N}$

$$\hat{\theta}_2^n = \frac{\hat{x} + \sigma - 2\sigma\theta_1^n s}{1 + 2\sigma}$$

We can calculate $\overline{\hat{\theta}_2}$ as

$$\overline{\hat{\theta}_2} = \frac{\hat{x} + \sigma - 2\sigma s\overline{\theta_1}}{1 + 2\sigma}$$

Substituting back to $\hat{x}$,

$$\hat{x} = 2\sigma t - \sigma + \frac{\hat{x} + \sigma - 2\sigma s\overline{\theta_1}}{1 + 2\sigma}$$
$$\hat{x} = \frac{1 + 2\sigma}{2\sigma}\left(2\sigma t - \sigma + \frac{\sigma}{1 + 2\sigma} - \frac{2\sigma}{1 + 2\sigma} s\overline{\theta_1}\right)$$

then back to $\hat{\theta}_2^n$:

$$\hat{\theta}_2^n = t - \frac{1}{2} + \frac{1}{2(1 + 2\sigma)} - \frac{s\overline{\theta_1}}{1 + 2\sigma} + \frac{\sigma}{1 + 2\sigma} - \frac{2\sigma\theta_1^n s}{1 + 2\sigma}$$
$$= t - s\left(\frac{1}{1 + 2\sigma}\overline{\theta_1} - \frac{2\sigma}{1 + 2\sigma}\theta_1^n\right)$$

The formulas for $N = 2$ trivially follows.

We note that **Corollary 1** also follows in the $N \geq 2$ general case, this can be seen with trivial algebra.

Analytical solution of the incentive-compatibility constraint:

$$\frac{1}{2}\left(\frac{k}{c} - \frac{(1 - 2\sigma)(\theta_2^{max} - t)}{\overline{\theta_1} + 2\sigma\theta_1^i} + \sqrt{\frac{p_j \Delta\theta_1}{\overline{\theta_1} + 2\sigma\theta_1^i}\left(\frac{k^2}{c^2} - \frac{(1 + 2\sigma)(\theta_2^{max} - t)^2}{\theta_1^i(\overline{\theta_1} + 2\sigma\theta_1^i)}\right)}\right)$$

Appendix A.5 Proof of Theorem 2

As explained in the main text, critical regulation level is described by equation 13:

$$s_i^{cri} = s_j^{cri} \qquad \text{(13 revisited)}$$

The analytical solution to this equation is

$$
\begin{aligned}
s_P^{cri} &= \frac{2c(1+2\sigma)(\theta_2^{max}-t)+k\Delta\theta_1-\sqrt{(2c(1+2\sigma)(\theta_2^{max}-t)+k\Delta\theta_1)^2-4kc(1+2\sigma)(\theta_2^{max}-t)\Delta\theta_1}}{2c\Delta\theta_1}\\
&= \frac{2c(1+2\sigma)(\theta_2^{max}-t)+k\Delta\theta_1-\sqrt{(2c(1+2\sigma)(\theta_2^{max}-t))^2+(k\Delta\theta_1)^2}}{2c\Delta\theta_2}\\
&= \left(\frac{(1+2\sigma)(\theta_2^{max}-t)}{\Delta\theta_1}+\frac{k}{2c}\right)-\sqrt{\left[\frac{(1+2\sigma)(\theta_2^{max}-t)}{\Delta\theta_1}\right]^2+\left[\frac{k}{2c}\right]^2}
\end{aligned}
$$

TODO: show Pareto-optimality

TODO: clarify intuitive criterion refinement

Appendix A.6 Proofs for Welfare Section

Appendix A.6.1 Receiver - Separating equilibrium

The receivers utility is t if she withdraws, 1 if the regime survives and she stays and 0 if the regime fails. The mass of agents who attack at a given fundamental θ is $\alpha(\theta)$. We denote by $\theta^\star$ the value of θ where $\alpha(\theta)=0$ and by $\theta_\star$ where $\alpha(\theta)=1$.With this notation the total payoff to all receivers is

$$
\mathbb{E}u = \underbrace{\int_{\theta_{min}}^{\theta_*} t\nu(\cdot)d\theta_2}_{\text{All withdraw}} + \underbrace{\int_{\theta_\star}^{\theta^\star} t\alpha(\theta)\nu(\cdot)d\theta_2}_{\alpha\text{ withdraw}} + \underbrace{\int_{\hat{\theta}}^{\theta^*} 1-\alpha(\theta)\nu(\cdot)d\theta_2}_{1-\alpha\text{ stay and survive}} + \underbrace{\int_{\theta^\star}^{\theta^{max}} 1\nu(\cdot)d\theta_2}_{\text{all stay and survive}}
$$

where for example in the benchmark case

$$
\alpha(\theta_2) = \frac{\hat{x}-\theta_2+\sigma}{2\sigma} = \frac{t(1+2\sigma)-\theta_1 s-\theta_2}{2\sigma}
$$

$$
\theta^\star = t(1+2\sigma)-\theta_1 s
$$

$$
\theta_\star = \hat{x}-\sigma = t(1+2\sigma)-\theta_1 s-2\sigma
$$

$$
\hat{\theta} = t-\theta_1 s
$$

$$
\hat{\theta}-\theta_\star = 2\sigma(1-t)
$$

$$
\theta^\star-\hat{\theta} = 2\sigma t
$$

$$
\alpha(\hat{\theta}) = t
$$

Now without algebra, graphically it is very easy to calculate

$$
\int_{\theta_\star}^{\hat{\theta}} \alpha(\theta) = (1-t^2)\sigma
$$

$$
\int_{\hat{\theta}}^{\theta^\star} \alpha(\theta) = t^2\sigma
$$

$$\int_{\theta_\star}^{\theta^\star} \alpha(\theta) = \sigma$$

so putting together

$$Eu = \frac{1}{2\eta}\left(t(\theta_\star - \theta_{min}) + \sigma t - t^2\sigma + (\theta^{max} - \hat{\theta})\right)$$

We can calculate

$$\frac{\partial Eu}{\partial s} = \frac{1}{2\eta}\left((1-t)\theta_1\right) > 0$$

this implies the receiver always prefers a higher intervention.

We can rewrite

$$Eu = \frac{1}{2\eta}\left((\theta^{max} - t\theta_{min}) - \sigma t(1-t) - (t - \theta_1 s)(1-t)\right)$$

All difference in payoff is captured by the second term. The first term is constant (only a function of exogenous parameters). So we can define the following simple measure of receiver's welfare, keeping only the endogenous variable (s) and the "interesting" exogenous parameters (σ, θ_1)

$$WR = (\theta_1 s - \sigma t)(1-t)$$

In particular, receiver's welfare in the least-costly separating equilibrium is

$$WR^{SE} = \left(p\theta_1^i s_i^{FB} + (1-p)\theta_1^j s^{cri} - \sigma t\right)(1-t)$$

while in pooling:

$$WR^{PE} = \overline{\theta}_1 s^P (1-t)$$

Appendix A.6.2 Receiver - Pooling equilibrium

Substituting pooling thresholds we obtain:

$$\alpha(\theta_2) = \frac{\hat{x} - \theta_2 + \sigma}{2\sigma} = \frac{t(1+2\sigma) - \overline{\theta}_1 s - \theta_2}{2\sigma}$$

$$\theta^\star = t(1+2\sigma) - \overline{\theta}_1 s$$

$$\theta_\star = \hat{x} - \sigma = t(1+2\sigma) - \overline{\theta}_1 s - 2\sigma$$

$$\hat{\theta}_i = t - \theta_1^i s - \frac{s p_j \Delta\theta_1}{1+2\sigma}$$

$$\hat{\theta}_j = t - \theta_1^j s + \frac{s p_i \Delta\theta_1}{1+2\sigma}$$

$$\hat{\theta}_i - \theta_\star = 2\sigma(1-t) + s(\overline{\theta}_1 - \theta_1^i) - \frac{s p_j \Delta\theta_1}{1+2\sigma} = 2\sigma(1-t) + s p_j \Delta\theta_1 \frac{2\sigma}{1+2\sigma}$$

$$\hat{\theta}_j - \theta_\star = 2\sigma(1-t) + s(\overline{\theta}_1 - \theta_1^j) + \frac{s p_i \Delta\theta_1}{1+2\sigma} = 2\sigma(1-t) - s p_i \Delta\theta_1 \frac{2\sigma}{1+2\sigma}$$

$$\theta^\star - \hat{\theta}_i = 2\sigma t + s(\theta_1^i - \overline{\theta}_1) + \frac{sp_j\Delta\theta_1}{1+2\sigma} = 2\sigma t - sp_j\Delta\theta_1\frac{2\sigma}{1+2\sigma}$$

$$\theta^\star - \hat{\theta}_j = 2\sigma t + s(\theta_1^j - \overline{\theta}_1) - \frac{sp_i\Delta\theta_1}{1+2\sigma} = 2\sigma t + sp_i\Delta\theta_1\frac{2\sigma}{1+2\sigma}$$

$$\alpha(\hat{\theta}_i) = t - \frac{sp_j\Delta\theta_1}{1+2\sigma}$$

$$\alpha(\hat{\theta}_j) = t + \frac{sp_i\Delta\theta_1}{1+2\sigma}$$

$$\int_{\hat{\theta}_i}^{\theta^\star} \alpha(\theta) = \sigma\left(t - \frac{sp_j\Delta\theta_1}{1+2\sigma}\right)^2$$

$$\int_{\theta_\star}^{\hat{\theta}_i} \alpha(\theta) = \sigma - \sigma\left(t - \frac{sp_j\Delta\theta_1}{1+2\sigma}\right)^2$$

$$\int_{\hat{\theta}_j}^{\theta^\star} \alpha(\theta) = \sigma\left(t + \frac{sp_i\Delta\theta_1}{1+2\sigma}\right)^2$$

$$\int_{\theta_\star}^{\hat{\theta}_j} \alpha(\theta) = \sigma - \sigma\left(t + \frac{sp_i\Delta\theta_1}{1+2\sigma}\right)^2$$

Substituting to the welfare function (consistent with bank-run story):

$$\mathbb{E}u^i = \frac{1}{2\eta}\left(t(\theta_\star - \theta_{min}) + t\sigma + (\theta^\star - \hat{\theta}) - \sigma(t - \frac{sp_j\Delta\theta_1}{1+2\sigma})^2 + (\theta^{max} - \theta^\star)\right)$$

$$= \frac{1}{2\eta}\left((t^{max} - t\theta_{min}) + (t^2(1+2\sigma) - t\overline{\theta}_1 s - 2\sigma t) + t\sigma - \sigma(t - \frac{sp_j\Delta\theta_1}{1+2\sigma})^2 - \hat{\theta}\right)$$

and

$$\mathbb{E}u^j = \frac{1}{2\eta}\left(t(\theta_\star - \theta_{min}) + t\sigma + (\theta^\star - \hat{\theta}) - \sigma(t - \frac{sp_j\Delta\theta_1}{1+2\sigma})^2 + (\theta^{max} - \theta^\star)\right)$$

$$= \frac{1}{2\eta}\left((t^{max} - t\theta_{min}) + (t^2(1+2\sigma) - t\overline{\theta}_2 s - 2\sigma t) + t\sigma - \sigma(t + \frac{sp_i\Delta\theta_1}{1+2\sigma})^2 - \hat{\theta}\right)$$

—

Appendix B Appendix B - Generalizations

Appendix B.1 Generalized functional forms

Model We leave the payoffs to the receivers and the gross payoff to sender unchanged. The generalization focuses on two components of the model:

(1) The cost of action is a generic function $c(s)$ with the assumptions $\frac{\partial c(s)}{\partial s} > 0$ and $\frac{\partial^2 c(s)}{\partial s^2} > 0$ that is the cost function is increasing and strictly concave in the signal s. This implies the net payoff for the sender is

$$U(\theta, \alpha) = \begin{cases} k - c(s) & \text{if} \quad \mathcal{R}(\theta_1, \theta_2, s, \alpha) \geq 0 \\ 0 & \text{if} \quad \mathcal{R}(\theta_1, \theta_2, s, \alpha) < 0 \end{cases}$$

(2) The regime change function is also a generic function $\mathcal{R}(\theta_1, \theta_2, s, \alpha)$ with the assumptions

$$\frac{\partial \mathcal{R}}{\partial \theta_1} > 0; \qquad \frac{\partial \mathcal{R}}{\partial \theta_2} > 0; \qquad \frac{\partial \mathcal{R}}{\partial s} > 0; \qquad \frac{\partial \mathcal{R}}{\partial \alpha} < 0; \qquad \frac{\partial^2 \mathcal{R}}{\partial \theta_1 \partial \theta_2} = 0; \qquad \frac{\partial^2 \mathcal{R}}{\partial s \partial \theta_1} > 0;$$

where the last condition plays the role of a single crossing condition. All other components of the model remain unchanged.

First-best The indifference condition remain unchanged since the generalization does not change directly the receivers' problem. The mass of agents who attack is therefore unchanged, so we conclude

$$\alpha = \frac{\hat{x} - \hat{\theta}_2 + \epsilon}{2\epsilon} = t$$

The failure condition solves after substituting $\alpha = t$

$$\mathcal{R}(\theta_1, \theta_2, s, t) = 0$$

from which [27]

$$\hat{\theta}_2 = f(t, \theta_1, s)$$

$$\hat{x} = f(t, \theta_1, s) + 2\sigma t - \sigma$$

We want to establish that the thresholds are decreasing in both θ_1 and s. This can be shown using *implifit function theorem* (IFT)

$$\frac{\partial \hat{\theta}_2}{\partial s} = -\frac{\frac{\partial \mathcal{R}}{\partial s}}{\frac{\partial \mathcal{R}}{\partial \theta_2}} = -\frac{[+]}{[+]} < 0$$

$$\frac{\partial \hat{\theta}_2}{\partial \theta_1} = -\frac{\frac{\partial \mathcal{R}}{\partial s}}{\frac{\partial \mathcal{R}}{\partial \theta_2}} = -\frac{[+]}{[+]} < 0$$

Which generalizes the result that higher signal as well as higher type decreases threshold, therefore improves stability and increases probability of survival of the regime.

[27] TBD: formal proof of the existence of $f(\cdot)$ is missing: after imposing assumptions on $\mathcal{R}$ which are required for dominance regions, then continuity and intermediate value theorem implies the existence.

Optimal policy

The expected profit for any given policy choice s is:

$$\pi(\theta_1, s) = \rho\,(k - c(s))$$

Optimum policy intervention is determined by

$$\frac{\partial \pi(\theta_1, s)}{\partial s} = (k - c(s))\,\frac{\partial \rho}{\partial s} - \rho \frac{\partial c(s)}{\partial s} = 0$$

but from the formula for $\rho = \frac{1}{2\eta}(\theta_2^{max} - \hat{\theta}_2)$

$$\frac{\partial \rho}{\partial s} = -\frac{1}{2\eta}\frac{\partial \hat{\theta}_2}{\partial s} > 0$$

so the FOC of optimality can be written as (to simplify notation: $\partial_x \mathcal{R} = \frac{\partial \mathcal{R}}{\partial x}$)

$$-\frac{K}{2\eta}\frac{\partial \hat{\theta}_2}{\partial s} = \frac{\partial c(s)}{\partial s}$$

$$\frac{K}{2\eta}\frac{\partial_s \mathcal{R}}{\partial_{\theta_2}\mathcal{R}} - \frac{\partial c(s)}{\partial s} = 0$$

The question is how $s^\star$ changes with θ_1. For that we use IFT on FOC. The second term doesn't change with θ_1. The first term

$$\partial \frac{\partial_s \mathcal{R}}{\partial_{\theta_2}\mathcal{R}} / \partial \theta_1 = \frac{\partial_{s\theta_1}\mathcal{R}\partial_{\theta_2}\mathcal{R} - \partial_{\theta_2\theta_1}\mathcal{R}\partial_s\mathcal{R}}{\partial^2_{\theta_2}\mathcal{R}}$$

$$\partial \frac{\partial_s \mathcal{R}}{\partial_{\theta_2}\mathcal{R}} / \partial s = \frac{\partial_{ss}\mathcal{R}\partial_{\theta_2}\mathcal{R} - \partial_{\theta_2 s}\mathcal{R}\partial_s\mathcal{R}}{\partial^2_{\theta_2}\mathcal{R}}$$

using assumptions $\partial_{\theta_2\theta_1}\mathcal{R} = 0$ and $\partial_{ss}\mathcal{R} \geq 0$ and $\partial_{\theta_2}\mathcal{R} > 0$ and $\partial_{\theta_2 s}\mathcal{R} = 0$ this implies

$$\text{sign}[\partial[RHS]/\partial\theta_1] = \text{sign}[\partial_{s\theta_1}\mathcal{R}] = [+]$$

$$\text{sign}[\partial[RHS]/\partial s] = \frac{[+]}{[+]}$$

$$\partial c(s)/\partial(s) \geq 0$$

another application of IFT implies for the case $\text{sign}[\partial_{\theta_2 s}] > 0$

$$\frac{\partial s^\star}{\partial \theta_1} = -\frac{\partial_{\theta_1} FOC}{\partial_s FOC} = -\frac{+}{[+] - [+]} > 0$$

as long as $\mathcal{R}$ is not too convex in s compared to the cost function. Precisely, it must be the case that

$$\frac{\partial_{ss}\mathcal{R}\partial_{\theta_2}\mathcal{R} - \partial_{\theta_2 s}\mathcal{R}\partial_s\mathcal{R}}{\partial^2_{\theta_2}\mathcal{R}} = \frac{\partial_{ss}\mathcal{R}}{\partial_{\theta_2}} < \frac{\partial c'(s)}{\partial s}$$

Appendix B.2 Separating equilibrium

Summary: For any arbitrary function $\mathcal{R}(\boldsymbol{\theta}, s, \alpha)$ we derive a first-order approximation of failure thresholds off-the-equilibrium path, i.e. when agents act as if the type is j while the actual type is i. Alternatively, this threshold can also be seen as a decomposition of the difference between equilibrium thresholds for type i and j into two components - a direct effect through regime change function, and an indirect effect which is purely belief-based. The main result is that as the noise becomes more precise ($\sigma \to 0$), the indirect effect dominates the direct effect, pushing the off-equilibrium threshold towards the other type's threshold. Intuitively, this increases the potential gains for the low type from mimicking the high type.

We know that

$$\alpha(\hat{x}, \theta_2) = \frac{\hat{x} - \theta_2 + \sigma}{2\sigma}$$

$$\hat{x} = \hat{\theta}_2 + 2\sigma t - \sigma$$

The relevant thresholds on- and off-equilibrium path respectively solve

$$\mathcal{R}(\theta_1^i, \theta_2, s^i, \alpha(\hat{x}^i, \theta_2)) = 0 \qquad [\hat{\theta}_2^i]$$

$$\mathcal{R}(\theta_1^i, \theta_2, s^j, \alpha(\hat{x}^j, \theta_2)) = 0 \qquad [\hat{\theta}_2^{i.j}]$$

$$\mathcal{R}(\theta_1^j, \theta_2, s^j, \alpha(\hat{x}^j, \theta_2)) = 0 \qquad [\hat{\theta}_2^j]$$

$$\mathcal{R}(\theta_1^j, \theta_2, s^i, \alpha(\hat{x}^i, \theta_2)) = 0 \qquad [\hat{\theta}_2^{j.i}]$$

Using implicit function theorem (IFT) we can calculate the derivative of the full-information fundamental threshold with respect to θ_1

$$\frac{\partial \hat{\theta}_2}{\partial \theta_1} = -\frac{\partial_{\theta_1}\mathcal{R}}{\partial_{\theta_2}\mathcal{R}} = \frac{\frac{\partial \mathcal{R}}{\partial \theta_1} + \frac{\partial \mathcal{R}}{\partial \alpha}\frac{\partial \alpha}{\partial \hat{x}}\frac{\partial \hat{x}}{\partial \hat{\theta}_2}\frac{\partial \hat{\theta}_2}{\partial \theta_1}}{\frac{\partial \mathcal{R}}{\partial \theta_2} + \frac{\partial \mathcal{R}}{\partial \alpha}\frac{\partial \alpha}{\partial \theta_2}}$$

This derivative can be used to approximate the change when type changes from θ_1^i to θ_1^j in 'full information' model, keeping signal s fixed. This defines a natural decomposition

$$\frac{\partial \hat{\theta}_2}{\partial \theta_1} = \underbrace{\frac{\frac{\partial \mathcal{R}}{\partial \theta_1}}{\frac{\partial \mathcal{R}}{\partial \theta_2} + \frac{\partial \mathcal{R}}{\partial \alpha}\frac{\partial \alpha}{\partial \theta_2}}}_{\text{Direct effect}} + \underbrace{\frac{\frac{\partial \mathcal{R}}{\partial \alpha}\frac{\partial \alpha}{\partial \hat{x}}\frac{\partial \hat{x}}{\partial \hat{\theta}_2}\frac{\partial \hat{\theta}_2}{\partial \theta_1}}{\frac{\partial \mathcal{R}}{\partial \theta_2} + \frac{\partial \mathcal{R}}{\partial \alpha}\frac{\partial \alpha}{\partial \theta_2}}}_{\text{Indirect effect}}$$

The difference between $\hat{\theta}_2^i$ and $\hat{\theta}_2^{i.j}$ is only through the second term, while the difference between $\hat{\theta}_2^{i.j}$ and $\hat{\theta}_2^j$ is only through the first term, which allows us to write

$$\hat{\theta}_2^{i.j} = \hat{\theta}_2^j + \frac{\frac{\partial \mathcal{R}}{\partial \theta_1}}{\frac{\partial \mathcal{R}}{\partial \theta_2} + \frac{\partial \mathcal{R}}{\partial \alpha}\frac{\partial \alpha}{\partial \theta_2}}\Delta\theta_1 = \hat{\theta}_2^j + \frac{\frac{\partial \mathcal{R}}{\partial \theta_1}}{\frac{\partial \mathcal{R}}{\partial \theta_2} - \frac{1}{2\sigma}\frac{\partial \mathcal{R}}{\partial \alpha}}\Delta\theta_1$$

$$\hat{\theta}_2^{i.j} = \hat{\theta}_2^i - \frac{\frac{\partial \mathcal{R}}{\partial \alpha}\frac{\partial \alpha}{\partial \hat{x}}\frac{\partial \hat{x}}{\partial \hat{\theta}_2}\frac{\partial \hat{\theta}_2}{\partial \theta_1}}{\frac{\partial \mathcal{R}}{\partial \theta_2} + \frac{\partial \mathcal{R}}{\partial \alpha}\frac{\partial \alpha}{\partial \theta_2}}\Delta\theta_1 = \hat{\theta}_2^i - \frac{\frac{1}{2\sigma}\frac{\partial \mathcal{R}}{\partial \alpha}\frac{\partial \hat{x}}{\partial \hat{\theta}_2}\frac{\partial \hat{\theta}_2}{\partial \theta_1}}{\frac{\partial \mathcal{R}}{\partial \theta_2} - \frac{1}{2\sigma}\frac{\partial \mathcal{R}}{\partial \alpha}}\Delta\theta_1$$

From here we can immediately generalize two results of the main text

(1) As $\sigma \to 0$ the direct effect goes to zero, thereby $\hat{\theta}_2^{i.j} \to \hat{\theta}_2^j$

(2) Due to assumptions on the derivatives, both effect terms are always positive, guaranteeing $\hat{\theta}_2^i \leq \hat{\theta}_2^{i.j} \leq \hat{\theta}_2^j$ (i and j interchangeable, $=$ only in limiting cases).

Comment for next step: when analysing the existence of SE, we only need to take into consideration the change in s, but not here, the analysis of this section is for fixed s.

Example Apply to example $\mathcal{R}$ the derivatives give precise solutions because of linearity.

$$\frac{\hat{\theta}_2}{\hat{\theta}_1} = \frac{s + \frac{1}{2\sigma}(s)}{1 + \frac{1}{2\sigma}} = s$$

which implies

$$\hat{\theta}_2^j - \hat{\theta}_2^i = s\Delta\theta_1$$

which is correct because from the formula $t - \theta_1^j s - t + \theta_1^i s = s\Delta\theta_1$.

The decomposition above implies

$$\hat{\theta}_2^{i.j} - \hat{\theta}_2^j = \frac{s}{1 + \frac{1}{2\sigma}}\Delta\theta_1 = \frac{2\sigma s}{1 + 2\sigma}\Delta\theta_1$$

and

$$\hat{\theta}_2^i - \hat{\theta}_2^{i.j} = \frac{\frac{1}{2\sigma}s}{1 + \frac{1}{2\sigma}}\Delta\theta_1 = \frac{s}{1 + 2\sigma}\Delta\theta_1$$

which is exactly what we have in the main text.

Appendix B.3 Pooling equilibrium

Equation 1 is unchanged:

$$\hat{x}^P = 2\sigma t - \sigma + p_i\hat{\theta}_2^i + p_j\hat{\theta}_2^j$$

Equation 2 can only be expressed implicitly:

$$\mathcal{R}(\theta_1, \theta_2, s, \alpha(\hat{x}, \theta_2)) = 0$$

Write the Taylor-approximation of the difference between $\hat{\theta}_2^i$ (full-information threshold) and $\hat{\theta}_2^{iP}$ (pooling threshold).

$$\frac{\partial\hat{\theta}_2}{\partial\hat{x}} = -\frac{\frac{\partial\mathcal{R}}{\partial\hat{x}}}{\frac{\partial\mathcal{R}}{\partial\theta_2}} = -\frac{\frac{\partial\mathcal{R}}{\partial\alpha}\frac{\partial\alpha}{\partial\hat{x}}}{\frac{\partial\mathcal{R}}{\partial\theta_2} + \frac{\partial\mathcal{R}}{\partial\alpha}\frac{\partial\alpha}{\partial\theta_2}} = -\frac{\frac{1}{2\sigma}\frac{\partial\mathcal{R}}{\partial\alpha}}{\frac{\partial\mathcal{R}}{\partial\theta_2} - \frac{1}{2\sigma}\frac{\partial\mathcal{R}}{\partial\alpha}} > 0$$

$$\hat{\theta}_2^{iP} = \hat{\theta_2}^i + \frac{\partial\hat{\theta}_2}{\partial\hat{x}}\Delta\hat{x}$$

where

$$\Delta\hat{x} = \hat{x}^P - \hat{x}^i = (1-p)\Delta\hat{\theta}_2 = -p_j s\Delta\theta_1 = -s(\overline{\theta}_1 - \theta_1^i)$$

The derivative goes to zero as $\frac{1}{\sigma} \to 0$. That implies, low precision pushes pooling thresholds (respectively, profits) towards the full-information thresholds. This is consistent with the 'smoothing out differences' intuition.

The results are symmetric for type j. First define

$$\Delta\hat{x}^j = \hat{x}^{jP} - \hat{x}^j = p_i\hat{\theta}_2^i - p_i\hat{\theta}_2^j = -p\Delta\hat{\theta}_2 = ps\Delta\theta_1 = -s(\overline{\theta}_1 - \theta_1^j)$$

then approximate the average thresholds

$$\overline{\theta}_2^P = p_i\theta_2^{iP} + p_j\theta_2^{jP} \cong \overline{\hat{\theta}_2} + \frac{\partial\hat{\theta}_2}{\partial\hat{x}}\overline{\Delta\hat{x}} = t - s\overline{\theta}_1 + \frac{\partial\hat{\theta}_2}{\partial\hat{x}}(-s)(\overline{\theta}_1 - \overline{\theta}_1) = t - s\overline{\theta}_1$$

This is again the same result as for the specific case, but this seem to hold generally!

Example

Using the example $\mathcal{R} = \theta_2 + \theta_1 s - \alpha = 0$ we have

$$\frac{\partial\hat{\theta}_2}{\partial\hat{x}} = \frac{1}{2\sigma + 1}$$

and

$$\hat{\theta}_2^{iP} = t - \frac{s\overline{\theta}_1}{1 + 2\sigma} - \frac{2\sigma}{1 + 2\sigma}s\theta_1^i$$

just as we have seen in the main text.

Appendix B.4 Regulation

Need to prove: there exists some level of s_p, such that $[IC_i]$ below is satisfied, but $[IC_j]$ is not. Interpretation: for sufficiently high s_p, the level of s which would prevent i from mimicking is not incentive-compatible for type j.

A candidate proof:

The incentive compatibility constraints for an arbitrary level of (minimum) pooling

$$\pi_{i.j}(s) \leq \pi_i(s_p) \qquad [IC_i]$$

$$\pi_j(s) \geq \pi_{j.i}(s_p) \qquad [IC_j]$$

These inequalities say that s must not be profitable for the low-type even if it is believed to be of high type, but must be profitable for the high-type, if otherwise is believed to be of low-type.

All 4 functions in the IC's reach zero at $s^{lim} = c^{-1}(k)$. Suppose our conjecture is correct. This is equivalent with the existence of some $s_p < s$ such that both inequalities are satisfied with equality at the same critical s, which just solves IC_i and just fail to solve IC_j. This critical situation, if exist, therefore characterized equivalently by a pair $\{s_p; s\}$ where $s_p < s$ which solves both inequality with equality. Consider all possible pairs of $\{s_p; s\}$, not just those which solves the IC's, and define the following profit-differences

$$\Delta\pi_i(s) = \pi_j(s) - \pi_{i.j}(s) := \varphi(s)$$

$$\Delta \pi_j(s) = \pi_{j.i}(s) - \pi_i(s) := \varphi(s)$$

Viewed as a function of an arbitrary s, these are both *the same* concave functions with zeroes at $\{0, s^{lim}\}$, which we denoted by $\varphi(s)$. If a critical pair $\{s_p; s\}$ exists, it must satisfy $\varphi(s_p) = \varphi(s)$, call this the 'critical condition'. In addition, if a candidate critical pair $\{s_p, s\}$ satisfies the critical condition and at least one IC, say $[IC_i]$, then automatically satisfies the other one. This follows from the definition of 'critical condition'.

We know that both $s_i^{cri}(s_p) \to s^{lim}$ from below, and $s_i^{cri}(s_p)$ increases in s_p and its image is a compact interval $[s_{i.2}^{cri}, s^{lim}]$. It is obvious to see that the solution of the critical condition (that is an explicit expression of pairs s_p, s viewed as a function), $s(s_p)$ is continuously decreasing in s_p with the image $[s^{max}, s^{lim}]$ where s^{max} is a value which maximizes $\varphi(s)$, the difference between profits. A sufficient condition for the existence of a critical s_p, s is that $s^{max} < s_{i.2}^{cri}$, for which, we only have to prove that $s_j^\star < s^{max}$. This is tivial (although some more formality would be nice here) as long as $\pi(s = 0) > \pi(s = s^{lim}) = 0$